

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 16:38:05 ; Search time 4703.97 Seconds
(without alignments)
10155.006 Million cell updates/sec

Title: US-09-596-141C-3

Perfect score: 1643

Sequence: 1 gaattcttctgtgtgctc.....caggcgagcgaggagctc 1643

Scoring table: IDENTITY_NUC

Gapop 10-C, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4-09280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank.*

1: gb_ba.*

2: gb_hug.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pi.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pi.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pln.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vrt.*

38: em_sy.*

39: em_hugo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1643	100.0	1643	6	AX060715	Sequence
2	1643	100.0	1643	6	AX060894	Sequence
3	1544.8	94.0	96717	9	AL359182	Human DNA
4	1488.6	91.2	183999	6	AX092589	Sequence
5	1485.4	90.4	3231	6	AX351029	Sequence
6	1418	86.3	149034	9	AF275948	Homo sapi
7	1394.4	84.9	2893	6	AX351031	Sequence
8	1247.6	75.9	201144	9	AF287262	Homo sapi
9	1007	61.3	175064	2	AC012230	Homo sapi
10	1006	61.2	1167	9	HA252201	Homo sapi
11	1002.8	61.0	1167	9	AF258623S1	Homo sapi
12	519.8	31.6	69570	2	AC021246	Homo sapi
13	271.6	15.5	69570	2	AC021246	Homo sapi
14	223.8	14.0	90698	2	AC021345	Homo sapi
15	173	10.5	697	9	AF258627	Homo sapi
16	155.4	9.5	186889	2	AL807243	Mus muscu
17	153.2	9.3	278572	10	AF287263	AF287263 Mus muscu
18	153.2	9.3	145833	2	AC125837	AC125837 Rattus no
19	102.2	6.2	2893	6	AX351031	Sequence
20	102.2	6.2	3231	6	AX351029	Sequence
21	99.4	6.0	56717	9	AL359182	Sequence
22	98	6.0	7260	6	AX253452	Human DNA
23	97.8	6.0	49034	9	AF275948	Sequence
24	97	5.9	1643	6	AX060715	Homo sapi
25	97	5.9	1643	6	AX060894	Sequence
26	92.6	5.6	183999	6	AX092589	Sequence
27	91	5.5	221	6	AX351032	Sequence
28	89.4	5.4	1556	9	AK024328	Sequence
29	89	5.4	1750	9	AK022254	Homo sapi
30	75	4.6	9854	6	AX127831	Sequence
31	75	4.6	9854	6	AX139818	Sequence
32	69.2	4.2	201144	9	AF287262	Homo sapi
33	67	4.1	10442	6	AX060713	Sequence
34	67	4.1	10442	6	AX060892	Sequence
35	67	4.1	10442	9	AF285167	Sequence
36	60	3.7	10474	6	AX060719	Homo sapi
37	60	3.7	10474	6	AX060721	Sequence
38	60	3.7	10474	6	AX060898	Sequence
39	60	3.7	10474	6	AX060900	Sequence
40	55.6	3.4	191720	2	AC109982	Sequence
41	52.6	3.2	125020	9	AF429315	Rattus no
42	52.2	3.2	125020	9	AF429315	Homo sapi
43	49.6	3.0	96136	2	AC097674	Homo sapi
44	49.4	3.0	173127	2	AC131142	Sequence
45	49	3.0	65780	2	AC111741	Rattus no

ALIGNMENTS

RESULT 1	AX060715	Sequence 3 from Patent WO0078972.	1643 bp	DNA	linear	PAT 22-JAN-2001
LOCUS	AX060715	Sequence 3 from Patent WO0078972.	1643 bp	DNA	linear	PAT 22-JAN-2001
DEFINITION	AX060715	Sequence 3 from Patent WO0078972.	1643 bp	DNA	linear	PAT 22-JAN-2001
ACCESSION	AX060715	Sequence 3 from Patent WO0078972.	1643 bp	DNA	linear	PAT 22-JAN-2001
VERSION	AX060715.1	Sequence 3 from Patent WO0078972.	1643 bp	DNA	linear	PAT 22-JAN-2001
KEYWORDS	human.	Sequence 3 from Patent WO0078972.	1643 bp	DNA	linear	PAT 22-JAN-2001
SOURCE	human.	Sequence 3 from Patent WO0078972.	1643 bp	DNA	linear	PAT 22-JAN-2001
ORGANISM	Homo sapiens	Sequence 3 from Patent WO0078972.	1643 bp	DNA	linear	PAT 22-JAN-2001
REFERENCE	1 (bases 1 to 1643)	Sequence 3 from Patent WO0078972.	1643 bp	DNA	linear	PAT 22-JAN-2001
AUTORS	Lawn,R.M., Wade,D. and Garvin,M.	Sequence 3 from Patent WO0078972.	1643 bp	DNA	linear	PAT 22-JAN-2001
TITLE	Regulation with binding cassette transporter protein abel	Sequence 3 from Patent WO0078972.	1643 bp	DNA	linear	PAT 22-JAN-2001
JOURNAL	Patent: WO 0078972-A 3 28-DEC-2000;	Sequence 3 from Patent WO0078972.	1643 bp	DNA	linear	PAT 22-JAN-2001

Pred. No. is the number of results predicted by chance to have a

CV THERAPEUTICS, INC. (US)
Location/Qualifiers
1. 1643
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 370 a 413 c 457 g 403 t
ORIGIN

Query Match 100.0%; Score 1643; DB 6; Length 1643;
Best Local Similarity 100.0%; P-vec. No. 0;
Matches 1643; Conservative 0; Mismatches 0; Gaps 0;

QY 1 GAATTCCTTCCTGGTGGCTCCACATGACCTTCAGGSCCTGGCTTCCTATGGGTC 60
DB 1 GAATTCCTTCCTGGTGGCTCCACATGACCTTCAGGSCCTGGCTTCCTATGGGTC 60
QY 61 TGTCCCTGAGTGTGATAGACCACTCAITGAGTACTGGCTTGGCTTGGCTGGCTGAGA 120
DB 61 TGTCCCTGAGTGTGATAGACCACTCAITGAGTACTGGCTTGGCTTGGCTGGCTGAGA 120
QY 121 TCCCTGTGACTGTAGCATGAGGGGGCTTGTGAGCTGAATGTCTGTATGAGGTGGTGG 180
DB 121 TCCCTGTGACTGTAGCATGAGGGGGCTTGTGAGCTGAATGTCTGTATGAGGTGGTGG 180
QY 181 AGTCTGGAAATATGATGGAGCTGGAGTGGGAAGAGATGAGCTTGGGGCACTCTCTC 240
DB 181 AGTCTGGAAATATGATGGAGCTGGAGTGGGAAGAGATGAGCTTGGGGCACTCTCTC 240
QY 241 ATGCCACCTCATCTGCGCAAACTCAGTCAAACTGTGAAGAGTCAATATGGAATCTG 300
DB 241 ATGCCACCTCATCTGCGCAAACTCAGTCAAACTGTGAAGAGTCAATATGGAATCTG 300
QY 301 CCCCTCAAGTGTGCTACAAAGGTATCTTGTCAAGTGTAGAGACCTTGTGGCTCCACGT 360
DB 301 CCCCTCAAGTGTGCTACAAAGGTATCTTGTCAAGTGTAGAGACCTTGTGGCTCCACGT 360
QY 361 GCACCTCCAGGGCTGCTGGGCTCTCTACGGGTCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 361 GCACCTCCAGGGCTGCTGGGCTCTCTACGGGTCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 421 CCCTCAGGGCAGATTAATATTAGACTCTCAGACTTGGACTTGGCTTGGCTGCTGCTGCT 480
DB 421 CCCTCAGGGCAGATTAATATTAGACTCTCAGACTTGGACTTGGCTTGGCTGCTGCTGCT 480
QY 481 AGTGTACATTTAGTGTGGCTGTGATGAATGACTTAATATTAATATTAATATTAATATTA 540
DB 481 AGTGTACATTTAGTGTGGCTGTGATGAATGACTTAATATTAATATTAATATTAATATTA 540
QY 541 GGCTGTGATTCCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
DB 541 GGCTGTGATTCCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
QY 601 CCCCTACGCCAAAGCAACAGATTAAGTGGAGTCTGGAGTGGCTGCTGCTGCTGCTGCTGCT 660
DB 601 CCCCTACGCCAAAGCAACAGATTAAGTGGAGTCTGGAGTGGCTGCTGCTGCTGCTGCTGCT 660
QY 661 ACAGTGCATATCTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
DB 661 ACAGTGCATATCTGCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
QY 721 TTTTATACAGGAGGCTGATCAATATTAATGAATTAAGAGGGGCTGCTGCTGCTGCTGCTGCT 780
DB 721 TTTTATACAGGAGGCTGATCAATATTAATGAATTAAGAGGGGCTGCTGCTGCTGCTGCTGCT 780
QY 781 TTCGTGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 781 TTCGTGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
QY 841 ATGAAGAGACAGTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
DB 841 ATGAAGAGACAGTAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
QY 901 TGGGAATCTCCAGGAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960

DB 901 TGGGAATCTCCAGGAGTGTAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 961 AAACAAAGACAGCCCAATTACCCAGAGGACTGTCGGCTTCCCTCACCACAGCCTAGGCC 1020
DB 961 AAACAAAGACAGCCCAATTACCCAGAGGACTGTCGGCTTCCCTCACCACAGCCTAGGCC 1020
QY 1021 TTTGAAGGAAACAAAGACAAAGCAAAATGATTGGCGTCCCTCAGGAGATTGAGCCTAG 1080
DB 1021 TTTGAAGGAAACAAAGACAAAGCAAAATGATTGGCGTCCCTCAGGAGATTGAGCCTAG 1080
QY 1081 AGCTCTCTTCCCTCCCAATCCCTCCCTCCGCTCAGGAAACTAACAAAGGAAAAAAATTT 1140
DB 1081 AGCTCTCTTCCCTCCCAATCCCTCCCTCCGCTCAGGAAACTAACAAAGGAAAAAAATTT 1140
QY 1141 GCGGAAGCAGGAGTTTAGAGGAAGCAAAATCCACTGGTGGCTTGGCTGCCGGGAGCTG 1200
DB 1141 GCGGAAGCAGGAGTTTAGAGGAAGCAAAATCCACTGGTGGCTTGGCTGCCGGGAGCTG 1200
QY 1201 GACTAGAGAGTCTGGCGGCGAGCCCGAGCCGCTCCCGCGCTCTTAGGCCGGCG 1260
DB 1201 GACTAGAGAGTCTGGCGGCGAGCCCGAGCCGCTCCCGCGCTCTTAGGCCGGCG 1260
QY 1261 GCGCCGCGGCGGAGGAGCGACACCGCCGAGCCCGCTTAAGACACCTGCTGTACCTCCA 1320
DB 1261 GCGCCGCGGCGGAGGAGCGACACCGCCGAGCCCGCTTAAGACACCTGCTGTACCTCCA 1320
QY 1321 CCGCCACCCACCCACCTCCCTCCCACTCCCTAGATGCTGCTGGCGGCTGAGACCTGCG 1380
DB 1321 CCGCCACCCACCCACCTCCCTCCCACTCCCTAGATGCTGCTGGCGGCTGAGACCTGCG 1380
QY 1381 CCGTTTAAAGGGCGGCGCCCGCTCCACCTGCTTCTGCTGAGTGAACCTACATAA 1440
DB 1381 CCGTTTAAAGGGCGGCGCCCGCTCCACCTGCTTCTGCTGAGTGAACCTACATAA 1440
QY 1441 CAGAGCCGGGAGCGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
DB 1441 CAGAGCCGGGAGCGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
QY 1501 TGGCTGCTGGTGCAGCCGAATCTATAAAGAACTAGTCCCGCAAAACCCCTTAATTGC 1560
DB 1501 TGGCTGCTGGTGCAGCCGAATCTATAAAGAACTAGTCCCGCAAAACCCCTTAATTGC 1560
QY 1561 GAGCAGAGTGTAGTGGGCGGAGCCCGAGAGCCCGAGAGCCCGAGAGCCCGAGAGCCCGAG 1620
DB 1561 GAGCAGAGTGTAGTGGGCGGAGCCCGAGAGCCCGAGAGCCCGAGAGCCCGAGAGCCCGAG 1620
QY 1621 CGGCAGGCGAGGGCGGAGGAGCTC 1643
DB 1621 CGGCAGGCGAGGGCGGAGGAGCTC 1643

RESULT 2
AX060894
LOCUS AX060894
DEFINITION Sequence 3 from Patent WO0078971.
ACCESSION AX060894
VERSION AX060894.1 GI:12406271
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1643)
AUTHORS Lawn, R.M., Wade, D., Oram, J.F. and Garvin, M.
TITLE ATP binding cassette transporter protein abcl polypeptides
JOURNAL Patent: WO 0078971-A 3 28-DEC-2000;
CV THERAPEUTICS, INC. (US)
FEATURES
Location/Qualifiers
source 1. 1643
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 370 a 413 c 457 g 403 t

chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep. This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HG9/Chr9>. RP11-2.7B7 is from the library RP11-11.1 constructed by the group of Pletzer de Jong. For further details see <http://www.chori.org/bacpac/home.htm>.

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-217B7. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true right end of clone RP11-217B7 is at 96717 in this sequence. The true left end of clone RP11-122F10 is at 72980 in this sequence. The true right end of clone RP11-31320 is at 2000 in this sequence.

FEATURES

```

source
    1..96717
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="9"
    /clone="RP11-217B7"
    /note="Sequence from AF275948 sequenced by National
    Institutes of Health, National Heart, Lung and Blood
    Institute, Bethesda, MD 20892, USA."
    34249..84273
    /note="Sequence from overlapping clone RP11-122F10
    (AC026643). Assembly confirmed by restriction digest."
    92050..92163
    /note="Sequence from reads from a short insert library
    derived from a single pUC clone. Restriction digest data
    confirm the assembly."
    92411..92557
    /note="Sequence from reads from a short insert library
    derived from a single pUC clone. Restriction digest data
    confirm the assembly."

```

BASE COUNT 27673 a 21138 c 20380 g 27526 t

ORIGIN

```

Query Match 94.0%; Score 1544.8; DB 9; Length 96717;
Best Local Similarity 99.4%; Pred. No. C;
Matches 1624; Conservative 0; Mismatches 2; Indels 8; Gaps 7;

QY 16 GGTCCACATGACCTCCAGGCGCTCTGG--CTCTTATGGTCTGNCCTGACTGTC 73
Db 18798 GCTCCACATGACCTCCAGGCGCTCTGGGCGCTCTTATGGTCTGNCCTGAGTGT 18739

QY 74 GATGAACCACTGATGTGAGTACTGGGCTTGAG-CGTGGCTGGAGATCCTGTGACTG 132
Db 18738 GATGAACCACTGATGTGAGTACTGGGCTTGAGCGCTGGCTGGAGATCCTGTGACTG 18679

QY 133 TAGCTGGAGGGGGCTTGT-CAGCTGAATGCTGTATGAGGTGGGGAGTTCIGGAAT 191
Db 18678 TAGCATGGAGGGGGCTGTGTACAGCTGAATGCTGTATGAGGTGGGGAGTTCIGGAAT 18619

QY 192 ATGATGGAGCTGGAGGTGGGAAGAGTGGCTTGGGGCAGCTCTCTCAIGCCACTCA 251
Db 18618 ATGATGGAGCTGGAGGTGGGAAGAGTGGCTTGGGGCAGCTCTCTCAIGCCACTCA 18559

QY 252 TTCTGGCCAAACTCAGGTCAAACCTGTGAAGAGTCTAAATGTGAATCTGCCCTTCAAGGT 311
Db 19558 TTCTGGCCAAACTCAGGTCAAACCTGTGAAGAGTCTAAATGTGAATCTGCCCTTCAAGGT 18499

```

```

QY 312 GGCTACAAAGGTATCTTTCTCAAGGTAGGAGACC-TGTGGCTCCACGTGCACCTCCAGG 371
Db 18498 GGCTACAAAGGTATCTTTCTCAAGGTAGGAGACC-TGTGGCTCCACGTGCACCTCCAGG 18439

QY 372 GCGTCTCTT-GGCGTCTCTACCGGCTGCTCCGAGCTCTTCTATGAATCT-CCTCTCAGG 429
Db 18438 GCGTCTCTTGGGCTCTCTCTACGGGCTGCTCCGAGCTCTTCTATGAATCTGCCCTTCAGG 18379

QY 430 CAGATTCATATTAGACTCTTCACAGTTTGACCTGAGTTTGGCCAGATAAGGTGACAT 489
Db 18378 CAGATTCATATTAGACTCTTCACAGTTTGACCTGAGTTTGGCCAGATAAGGTGACAT 18319

QY 490 TTAGTTTGTGGCTTGATGAATGACTTAAATATTAGACATATGGTGCTAGGCGCTGCAT 549
Db 18318 TTAGTTTGTGGCTTGATGAATGACTTAAATATTAGACATATGGTGCTAGGCGCTGCAT 18259

QY 550 TCCACTCTTGCTCTTTTGTGGCCCTCCAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 509
Db 18258 TCCACTCTTGCTCTTTTGTGGCCCTCCAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 18200

QY 610 GCCAAAGGCAACAGATAAGTTGGAGTCTGGAGTGGCTACATATATTTACACGACTGCA 669
Db 18199 GCCAAAGGCAACAGATAAGTTGGAGTCTGGAGTGGCTACATATATTTACACGACTGCA 18140

QY 670 ATTCTGTGGCTGCACCTTCACAAATGATATACAACTAAATACAAAG-CCTCTGTTTATCA 729
Db 18139 ATTCTGTGGCTGCACCTTCACAAATGATATACAACTAAATACAAAGTCTCTGTTTATCA 18080

QY 730 CAGGAGGCTGATCAATATAAATAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 789
Db 18079 CAGGAGGCTGATCAATATAAATAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 18021

QY 790 TTGTTTGTGGTGTCTCTTTTGTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 849
Db 18020 TTGTTTGTGGTGTCTCTTTTGTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 17961

QY 850 AGCAGTAAGATGTTCTCTCGGCTCCTCTGAGGAGCTGGGAGCTCAGCTGGGAGTCT 909
Db 17960 AGCAGTAAGATGTTCTCTCGGCTCCTCTGAGGAGCTGGGAGCTCAGCTGGGAGTCT 17901

QY 910 CCAAGCAGTAGTTCGCTCTCAAAATCAAAGTCCAGTTCTGGGGGAGGAGGAGGAGGAG 969
Db 17900 CCAAGCAGTAGTTCGCTCTCAAAATCAAAGTCCAGTTCTGGGGGAGGAGGAGGAGGAG 17841

QY 970 CAGCCCATACCCAGAGGACTGTCCGCTTCCCTCACCACGCTAGGCTTGAAGG 1029
Db 17840 CAGCCCATACCCAGAGGACTGTCCGCTTCCCTCACCACGCTAGGCTTGAAGG 17781

QY 1030 AACAAAAGCAAGACAAATGATTTGGCTGCTGAGGAGATTCAGCTAGACTCTCTC 1089
Db 17780 AACAAAAGCAAGACAAATGATTTGGCTGCTGAGGAGATTCAGCTAGACTCTCTC 17721

QY 1090 TCCCCCAATCCCTCCGCTGAGGAACTAAACAAAGGAGGAGGAGGAGGAGGAGGAGGAGG 1149
Db 17720 TCCCCCAATCCCTCCGCTGAGGAACTAAACAAAGGAGGAGGAGGAGGAGGAGGAGG 17661

QY 1150 AGGATTTAGAGGAGCAAAATTCACATGGTGGCTTGGCTGCCGGGAGCTGGAGTASAGA 1209
Db 17660 AGGATTTAGAGGAGCAAAATTCACATGGTGGCTTGGCTGCCGGGAGCTGGAGTASAGA 17601

QY 1210 GTCTGGGGCAGCCCGGAGCCCGAGGCTTCCGCGGCTTTAGGCGGCGGCGGCGGCGG 1269
Db 17600 GTCTGGGGCAGCCCGGAGCCCGAGGCTTCCGCGGCTTTAGGCGGCGGCGGCGGCGGCGG 17541

QY 1270 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1329
Db 17540 GGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 17481

QY 1330 CACCCACCTCCCCCAACTCCCTAGATGTGTCTGGGCGGCTGAACGTGCGCGCTTAAG 1389
Db 17480 CACCCACCTCCCCCAACTCCCTAGATGTGTCTGGGCGGCTGAACGTGCGCGCTTAAG 17421

```


Tue Feb 4 09:39:37 2003

DQ 2838 CCTCTGGCGTGGGTCAGCGCAATCTATAAAGAAGTAAGTCCGGGAAAAACCCCGTAA 2897

QY 1557 TTGCAGCGAGATGAGTTGGGCGCGGAGACCAGAGAGCCGAGCGGACGCTTCTCTCCCGG 1616
|||||
DQ 2898 TTGCAGCGAGATGAGTTGGGCGCGGAGACCAGAGAGCCGAGCGGACGCTTCTCTCCCGG 2957
|||||

QY 1617 GCTCAGCGAGCGAGCGGCGGAGCTC 1643
|||||

DQ 2958 GCTCAGCGAGCGAGCGGCGGAGCTC 2984
|||||

RESULT 6
AF275948

LOCUS AP275948 149034 bp DNA linear PRI 17-JUL-2000

DEFINITION Homo sapiens ASCA1 (ABCA1) gene, complete cds.

ACCESSION AF275948

VERSION AF275948.1 GI:9247085

KEYWORDS .

SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 149034)
Santamarina-Fojo,S., Peterson,K., Knapper,C., Qiu,Y., Freeman,L.,
Cheng,J.F., Osorio,J., Renaley,A., Yang,X.P., Haudenschild,C.,
Prades,C., Chimini,G., Blackmon,E., Francois,T., Duverger,N.,
Rubin,E.M., Rosier,M., Benefie,P., Fredrickson,D.S. and Brewer,H.B.
Jr.
TITLE Complete genomic sequence of the human ABCA1 gene: analysis of the
human and mouse ATP-binding cassette A promoter
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (14), 7587-7592 (2000)

PUBMED 20345099

REFID 10384428

AUTHORS 2 (bases 1 to 149034)
Santamarina-Fojo,S., Peterson,K.M., Knapper,C.E., Freeman,L.A.,
Renaley,A.T., Yang,X.-P., Haudenschild,C.C., Blackmon,E.E.,
Francois,T.L. and Brewer,H.B. Jr.
Direct Submission
Submitted (08-JUN-2000) Molecular Disease Branch, National
Institutes Of Health, National Heart, Lung and Blood Institute,
Bethesda, MD 20892, USA
Location/Qualifiers
1..149034
/organism="Homo sapiens"
/db_xref="taxon:9606"
1454..148034
/gene="ABCA1"

FEATURES source
gene
mRNA
join(1454..1674,25831..25989,40385..40478,45012..45153,
46423..46541,67729..67850,70831..71007,83960..84052,
89010..89250,91963..92101,92433..92549,96758..96955,
97702..97907,98428..98504,100391..100613,102360..102581,
103642..103846,104951..105064,106862..107033,
108023..108154,109450..109602,109806..109843,
110646..110866,112124..112196,113183..113385,
115053..115101,115297..115410,116806..116954,
118504..118726,119964..120062,123094..123283,
124804..124998,126208..126240,127363..127466,
128943..129017,129539..129708,130939..131118,
133114..133229,133341..133485,134526..134649,
135736..135865,136130..136250,137038..137100,
138008..138114,140469..140610,140982..141116,
142060..142163,142646..142738,143397..143640,
144581..148034)
/gene="ABCA1"
/product="ABCA1"
gln(1454..1674,25831..25922)
/gene="ABCA1"
2406..2678
/rpt_family="Alu"
repeat_region 4240..4509
/rpt_family="Alu"
repeat_region 4968..5270
/rpt_family="Alu"

TFVLELFTDNKLNINILKSVLIFPHFCIGRGLIDMVKNQAMADALERSGENREVS
PLSWDLVRNLFAMAVEGWFFLITVLIORFFIRPRVNAKLSPLANDEDDSVRRERQ
RILDGGONDILEIKELIKYRRKRKPAVDRIQVIPPGEFCGLIGVNGAKSSFTFKM
LTGDTVTIRGDAFNKXNLSILNHEVHQNMGYCPOFDAITELLTGREHVEFFALLRGV
PEKVKYGEWAKRLGLVKYGEKVAGNYSGRKRLSTAMALIGGPVVFIDEPTTG
MDPKRRYLWNAICLSVYKESGSLVTSMSRECEALCTEMA IWNWRFRCLGSOVHLK
NRFGGYITVIAAGSNDDLPVQDFGLAPGVSXKXKRNMLQYQLPSS:SSLARI
PSILSOKKRRIEDYSVQTTLDQVFNFAKDDSDHLKLSLHKNTQVVDVAVLI
SPQDEKRESVY
repeat_region 29916..30180
/rpt_family="Alu"
repeat_region 31050..31314
/rpt_family="Alu"
repeat_region 31692..31973
/rpt_family="Alu"
repeat_region 36167..36440
/rpt_family="Alu"
repeat_region 37617..37892
/rpt_family="Alu"
repeat_region 37999..38166
/rpt_family="Alu"
repeat_region 38172..38269
/rpt_family="Alu"
repeat_region 38995..39271
/rpt_family="Alu"
STS 41547..42585
/gene="ABCA1"
/note="A009X28"
repeat_region 41593..41835
/rpt_family="Alu"
STS 41657..41679
/gene="ABCA1"
/note="A009X28"
repeat_region 42320..42584
/rpt_family="Alu"
repeat_region 44349..44626
/rpt_family="Alu"
repeat_region 48473..48669
/rpt_family="Alu"
repeat_region 50547..50614
/rpt_family="Alu"
repeat_region 54385..54418
/rpt_family="Alu"
repeat_region 54849..54988
/rpt_family="Alu"
repeat_region 55029..55226
/rpt_family="Alu"
repeat_region 55297..55344
/rpt_family="Alu"
repeat_region 58501..58935
/note="LTR"
repeat_region 61246..61489
/rpt_family="Alu"
repeat_region 62812..63077
/rpt_family="Alu"
repeat_region 64382..64306
/rpt_family="Alu"
repeat_region 67123..67382
/rpt_family="Alu"
repeat_region 68499..68731
/rpt_family="Alu"
repeat_region 69481..69760
/rpt_family="Alu"
repeat_region 71996..72233
/rpt_family="Alu"
repeat_region 72244..72501

Query Match 86.3%; Score 1418; DB 9; Length 149034;
Sest Local Similarity 98.3%; Pred. No. 0;
Matches 1522; Conservative 0; Mismatches 5; Indels 21; Gaps 8;
QY 113 CTTGGAGATCCITGTTACATGTAGCATGGAGGGGGCTTGT-CAGCTGAATGTCGTATGCA 171

Db 1 CTTGGAGATCCITGTTACATGTAGCATGGAGGGGGCTTGTGCACTGAATGTCGTATGCA 60
QY 172 GGTGGTGGAGTCTTGGATATATGATGGACTGGAGGIGGAGAGAGAGTAGCTTGGGGC 231
Db 61 GGTGGTGGAGTCTTGGATATATGATGGACTGGAGGIGGAGAGAGAGTAGCTTGGGGC 120
QY 232 AGTCTCTCATGCACCTCATTCCTGGCCAAAACCTCAGGTCAAACTGTGAAGAGTCTTAAT 291
Db 121 AGTCTCTCATGCACCTCATTCCTGGCCAAAACCTCAGGTCAAACTGTGAAGAGTCTTAAT 180
QY 292 GTGAATCTGCCCTTCAAGGTGGCTACAAAGGTATCTTTGTCAAGGTAGGAGACCTTGTGG 351
Db 181 GTGAATCTGCCCTTCAAGGTGGCTACAAAGGTATCTTTGTCAAGGTAGGAGACCTTGTGG 240
QY 352 CTTCCAGGTGCACCTCCAGGGGCTGCTT-GGCTCTCTTACGGGCTGTGCTCAGTCTTC 410
Db 241 CTTCCAGGTGCACCTCCAGGGGCTGCTTGGGGCTCTCTTACGGGCTGCTCAGTCTTC 300
QY 411 TATGAATCT-CCCTTCAGGGCAGATTCATATTTAGACTCTTCACASTTTCACTTGAGTTT 469
Db 301 TATGAATCTCTCTTCAGGGCAGATTCATATTTAGACTCTTCACASTTTCACTTGAGTTT 360
QY 470 TGGCCAGAAATAGGTGACATTTAGTTTGGCTTGAATGACATTAATATATAGACA 529
Db 361 TGGCCAGAAATAGGTGACATTTAGTTTGGCTTGAATGACATTAATATATAGACA 420
QY 530 TATGGTGTGTAGGCTGCATTCCTACTCTGCTCTTTTGTGCTCTTACGGGCTTGTGG 589
Db 421 -TGGTGTGTAGGCTGCATTCCTACTCTGCTCTTTTGTGCTCTTACGGGCTTGTGG 478
QY 590 GTAGTTTGTGCTCCCTTACAGCCAAAGCAACAGATAAGTTGGAGTCTGGAGTGGCTA 649
Db 479 GTAGTTTGTGCT-CCCTTACAGCCAAAGCAACAGATAAGTTGGAGTCTGGAGTGGCTA 537
QY 650 CATAATTTTACAGACTGCAATTCCTGCTGCTCAAAATGTAATGAATTAATAATAATA 709
Db 538 CATAATTTTACAGACTGCAATTCCTGCTGCTCAAAATGTAATGAATTAATAATAATA 597
QY 710 CAAGTCTGTGCTTTTATCAGGAGGCTGATCAATAATTAATAATAATAATAATAATA 769
Db 598 CAAGTCTGTGCTTTTATCAGGAGGCTGATCAATAATTAATAATAATAATAATAATA 657
QY 770 GTCCCATATTCCTGCTGCTTTTGTGCTTTTGTGCTTTTGTGCTTTTGTGCTTTTGT 820
Db 658 GT-CCATATTTGCTGCTTTTGTGCTTTTGTGCTTTTGTGCTTTTGTGCTTTTGTG 716
QY 821 GGCTCTCTCTCTCAATTTATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 880
Db 717 GGCTCTCTCTCTCAATTTATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 776
QY 881 GGGACCTGGGAGCTCAGGCTGGGAATCTCCAAGGAGTGTGCTCCCTATCAAAATCAA 940
Db 777 GGGACCTGGGAGCTCAGGCTGGGAATCTCCAAGGAGTGTGCTCCCTATCAAAATCAA 836
QY 941 AGTCAGGTTTCTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1000
Db 837 AGTCAGGTTTCTGGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 896
QY 1001 CCTCAGCCCTCAGGCTTAGGCTTTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1060
Db 897 CCTCAGCCCTCAGGCTTAGGCTTTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 956
QY 1061 CTGAGGAGATTCAGGCTTAGGCTCTCTCTCCCAATCCCTCCCTCCCTCCCTCCCT 1120
Db 957 CTGAGGAGATTCAGGCTTAGGCTCTCTCTCCCAATCCCTCCCTCCCTCCCTCCCT 1016
QY 1121 TAACAAAG 1180
Db 1017 TAACAAAG 1076
QY 1181 CTTTGGCTCCGGGACGTGGAGTGTGGAGTGTGGAGTGTGGAGTGTGGAGTGTGGAGT 1240


```

join(58411, 58476, 72893, 72986, 77520, 77661, 78931, 79040)
101048, 101169, 101152, 104328, 117287, 117375,
2329, 122569, 122285, 123424, 125754, 125870,
130079, 130276, 131023, 131228, 131749, 131925,
133708, 133930, 135677, 135898, 136959, 137163,
138268, 138381, 140139, 140350, 141340, 144741,
142777, 142919, 143123, 143260, 143963, 144183,
145441, 145513, 145657, 146859, 148527, 148575,
148771, 148884, 150280, 150428, 152078, 152302,
153438, 153536, 150588, 156757, 158278, 158372,
153682, 159714, 160837, 160942, 162417, 162491,
163013, 163182, 164413, 164590, 165888, 165993,
166815, 166959, 169505, 168628, 169715, 169844,
170109, 170229, 171017, 171079, 171987, 172093,
174448, 174589, 174961, 175095, 176039, 176142,
176625, 176717, 177376, 177619, 178560, 178700)
/gene="ABCA1"
/codon_start=1
/product="ATP-binding cassette 1 sub-family A member 1"
/protein_id="AAK4326.1"
/db_xref="GI:13876613"
/translation="MACWPOLRLAKNLTFRROTCOLLLEVAPEGLFIFLIISVRL
SYPYEHCCEFPNKAQSGPTPWVGILICRNNPCFPRPGAPGVGNFNKSIY
ARLPSDARRLLYSOKDTSMKRWKRLQLQKKSSNLAKQDLVNEIFSGFLYH
NLSPLKIPMKRDAVILHKLFLQYLHSLNGSKSSNLAKQDLVNEIFSGFLYH
KEKLAAERVLNRMDILKILRTLSTSPFSGKELAEATKILSHLDEQLAEQIFSMR
GWSGDOVWFLFNVAQSSSSSTCIYQAAVRISVCGHPDEGGKLIKSLNVEENNYKALF
SGNCTEDATFYDNTSPYCNIDMKNLSSRLPIIKWALKLVXGILVLTPTPTKAF
POVAEYNYKTOLEAVPHDLSEWKEELSPKIFTEMENSOEDMLRMLDLSNDRHPWE
QQLDGLDWQAQDIVAPLAKHPEDVQSSNSYVTPRPAENETNOAIRTISRMPCWHLN
KLEPIATEVLNKLSEMLDERKFAGIVFTGLIGESILPHVHKYKIMQIDAVERT
NIKDGWDYDPRADPEMDRYVWGFGVADQVDEWQALIRVLTKTKTGIVQMCPY
PCYDQVITLRVMSRMLPTLAWIYSVAVILKGVIEKAEKALTMRIDGLNLSILW
7'SWTISELLPVLGAGLVLVILKGLNLTSPDSVYVFLSFVAVVTILQCLTILF
SRNLAAACGGLIYFTVLLPYVLVCVAMQDYVGTGLKIFASLLSPAGFGEGCFALFE
PQGCGVOMNDLPFSEVDEGPNLTTSKVSMLNFTPLVGMVMTIDAVTFPGVQGIPTPW
YPTCKLWFRSEDESKSHPSGNSKRLSEICMEEBPTHKLGVSINLKVYSDGMKV
AVDGLALNFDEQITSLPHNGAGKTTTMSILTGLPPTSGTAYILGVDISEKSTIR
ONLVCYPOHNYLFMLTVEBHHVYAHKLGSKKHVKAEQALDGLPSKSLASKSI
SQLSGMKRKLFLVALFVGGSKVYLDEPTAGDYVSRAGWELLKLYRQGTILST
HMRMDPVLGRIAIISHGKICVGSLLFKNGLTGYXTILVKKDESSLSSCRNS
STCVLKEDSVSSQSDAGLGSOHESDPTIDVSAISNLIRKHSEARLVEDIGHEL
TVLVPRAKGAQVFLPHETIDRLDGLISYSGISETWLEETFLKVAEESGVDAETS
ZCTPARSNRAGKQSCGLAPPTEDDAAAPNSDIDPESRETLLSGMDGKGYQVK
GKLTLOQFVALLKRLLIARSKGPFAOIVLPAVTPCVLAFVSLVPPCKYPSLE
LOPMYNEQITVSNADPEDTGIILELNLITKDPGTGTCMGENPITDPTCOAGBEW
TATAPYQITMDLFONGNWNQNSPACQSDSKTKMLPCCPAGGCTPPQKQNTA
DLDQLTGRNTSDLYATYQIIASLKLNKIWNEPRYGFSGUGVSNLQALPPSQEVN
DAIKOMKRLKLDSDADRIILSGREMTCLTNRNKKVFNFNKNGHAIISFLFAN
NAIIRANQKQENSHVIGITAFNEPZNLITQOLESALMTYDVLISVCITPAMSFV
PASVFTVLPQRVSKAKHLQITSGVKVPLVSLNFMVMDNVMVPAVLTIIITFCPQ
KSYVSTNLPVALLLILYKGSITPLFPHYKPAFVPISTAVTLYSNLFIIGNSVA
TVLDELFDLNNITDILKSVLFIHPHCLGRLIDMKVQKMAADALPFGENRVS
PLUSDVGRNFAMAVEGWFELTVLIQYRFTIRPPRVNAKLSPLNDEQDVRERQ
RIIDGGQNDILETKELTKIYIRKRPKPAVDICVGPTEGFCGLUGVAGKSTFKM
LTGDTITVRGDAFLNKLNSNIHRIENHMGYQCFDAITELTIGREYVEGSLFLRGY
PKQKYGKWEAIKRLGVLGKGYKAGYSGNKGKRLSTAMALIGGVPVFLDEPTGK
NDPKARBPGLWALSVLRGKSVYLTSHSCECALCTRAMALIGGVPFCGLGVSQHLK
NRFQGYITVIRIAGSNEDLAPVDQFPLAPGCVSLKEKRNLMLOYCLPSLSLARI
FTLSOSKRLKESYV
72893, 72986
/gene="ABCA1"
/number=3
77520, 77661
/gene="ABCA1"
/number=4
78931, 79049
/gene="ABCA1"
/number=5
101048, 101169
/gene="ABCA1"
/number=6

```


exon 104152..104328
/gene="ABCA1"
/number=7
exon 117287..117379
/gene="ABCA1"
/number=8
exon 122329..122569
/gene="ABCA1"
/number=9
exon 125285..125424
/gene="ABCA1"
/number=10
exon 125754..125870
/gene="ABCA1"
/number=11
exon 130079..130276
/gene="ABCA1"
/number=12
exon 131023..131228
/gene="ABCA1"
/number=13
exon 131749..131925
/gene="ABCA1"
/number=14
exon 133708..133930
/gene="ABCA1"
/number=15
exon 135677..135898
/gene="ABCA1"
/number=16
exon 136959..137163
/gene="ABCA1"
/number=17
exon 138268..138381
/gene="ABCA1"
/number=18
exon 140175..140350
/gene="ABCA1"
/number=19
exon 141340..141471
/gene="ABCA1"
/number=20
exon 142777..142919
/gene="ABCA1"
/number=21
exon 143123..143260
/gene="ABCA1"
/number=22
exon 143963..144183
/gene="ABCA1"
/number=23
exon 145441..145513
/gene="ABCA1"
/number=24
exon 146657..146859
/gene="ABCA1"
/number=25
exon 148527..148575
/gene="ABCA1"
/number=26
exon 148771..148884
/gene="ABCA1"
/number=27
exon 150280..150428
/gene="ABCA1"
/number=28
exon 152079..152202
/gene="ABCA1"
/number=29
exon 153438..153536
/gene="ABCA1"
/number=30
exon 156368..156757

Query Match 75.9%; Score 1247.6; DB 9; Length 201144;
Best Local Similarity 88.9%; Pred. No. 0;
Matches 1464; Conservative 0; Mismatches 129; Indels 54; Gaps 10;
/gene="ABCA1"
QY 16 GGCCTCCACATGCACTCCAGGGCCTGCTGG--CTCTTCATATGGGTCTGCTCAGTGT 73
DB 32410 GCCTCCACATGCACTCCAGGGCCTGCTGGCTCTCTAATGGTCTGCTCAGTGT 32469
QY 74 GATAGAACCACTGAATGAGTACCTGGCTTGAG--CGTGGCTGGAGATCCTGTGACTG 132
DB 32470 GATAGAACCACTGAATGAGTACCTGGCTTGAGCCGCTGGCTGGAGATCCTGTGACTG 32529
QY 133 TAGCATGAGGGGGCTTGT--CAGCTGAATGCTGTATGACAGTGGTGGAGTCTGGGAAT 191
DB 32530 TAGCATGAGGGGGCTTGTGACAGTGAATGCTGATCAGTGGTGGAGTCTGGGAAT 32589
QY 152 ATGATGAGCTGGAGTGGAGAGAACTAGGCTTGGGGCAGCTCTCATGCCACCTCA 251
DB 32590 ATGATGAGCTGGAGTGGAGAGAACTAGGCTTGGGGCAGCTCTCATGCCACCTCA 32649
QY 252 TTCTGGCCAAACTCAGGTCACAACTGTGAAGAGICTAAATGTGAATCTGCCCTTCAAGT 311
DB 32650 TTCTGGCCAAACTCAGGTCACAACTGTGAAGAGTCTAAATGTGAATCTGCCCTTCAAGT 32709
QY 312 GGCTACAAAGTATCTTTGTCAAGGTAGGAGACCTTGTGGCTCCACAGTGCATCCAGG 371
DB 32710 GGCTACAAAGTATCTTTGTCAAGGTAGGAGACCTTGTGGCTCCACAGTGCATCCAGG 32769
QY 372 GCGTGCCT--GGCCTCTTACGGCTCTGCTCAGTCTCTATGATCT--CCTTTCAGGG 429
DB 32770 GCGTGCCTCTTACGGCTCTGCTCAGTCTCTATGATCTCTATGATCTCTGATGAT 32829
QY 430 CAGATTCTATTTAGACTCTTCCACAGTGTGACCTGAGTGTGGCCAGAAATAGGTGACAT 489
DB 32830 CAGATTCTATTTAGACTCTTCCACAGGGGATCAATAGGTGCCCATCATGCTGCTG 32883
QY 490 TTAGTTTGTGGCTTGATGAATGACTTAAATATTTAGACATATGGTGTAGGCTTGCAT 549
DB 32884 -----ACTCAGTNNNAGNCCCTCTCCNNNNNNNN 32916
QY 550 TCCTACTCTTGGCTTTTGGCCCTCCAGTGTGTTGGTAGTGTGCTGCCCTTACA 609
DB 32917 NNNNNNNNNCCCTTTTGGCCCTCCAGTGTGTTGGTAGTGTGCTGCCCTTACA 32975
QY 610 GCCAAGGCAACAGATAAGTTGGAGTCTGGAGTGTGCTACATAAATTTACACGACTGCA 669
DB 32976 GCCAAGGCAACAGATAAGTTGGAGTCTGGAGTGTGCTACATAAATTTACACGACTGCA 33035
QY 670 ATTCTCTGGCTGCACCTCACAAATGTATACAACTAAATACAGTCTGTGTTTATCA 729
DB 33036 ATTCTCTGGCTGCACCTCACAAATGTATACAACTAAATACAGTCTGTGTTTATCA 33095
QY 730 CAGGAGGCTGATCAATATATGAAATTAAGGAGGGGTGGTCCCATATGCTCTGTGTT 789
DB 33096 CAGGAGGCTGATCAATATATGAAATTAAGGAGGGGTGGTCCCATATGCTCTGTGTT 33154
QY 790 TTTGTTGTTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 841
DB 33155 TTTGTTGTTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 33214
QY 842 TGAAGAGAACGAGTAAGATGTCCTCGGTCCTCTGAGGAGCTGGGAGCTCAGCT 901
DB 33215 TGAAGAGAACGAGTAAGATGTCCTCGGTCCTCTGAGGAGCTGGGAGCTCAGCT 33274
QY 902 GGAATCTCCAGGCGAGTAGTCCGCTATCAAAATCAAAATCAAAATCAAAATCAAAAT 961
DB 33275 GGAATCTCCAGGCGAGTAGTCCGCTATCAAAATCAAAATCAAAATCAAAATCAAAAT 33334
QY 962 ACACAAAGCAGCCCATTTACCCAGAGGACTGTCCGCTTCCCTTCCCTTCCCTTCCCTTCCCT 1021
DB 33335 ACACAAAGCAGCCCATTTACCCAGAGGACTGTCCGCTTCCCTTCCCTTCCCTTCCCTTCCCT 33394

```

QY 1022 TTGAAGAGAAACAAACAGACAAATGATTGGGTCCTGTAGGAGAGATTCAGCCTAGA 1081
DB 33395 TTGAAGAGAAACAAACAGACAAATGATTGGGTCCTGTAGGAGAGATTCAGCCTAGA 33454
QY 1082 GCTCTCTCTCCCAATCCCTCCGCTGAGGAACCTAACAAGAGAAAAAATTCG 1141
DB 33455 GCTCTCTCTCCCAATCCCTCCGCTGAGGAACCTAACAAGAGAAAAAATTCG 33514
QY 1142 CGGAAGACAGATTAGAGGAAGCAAAATCCACTGTGTCCTTGGCTGCGGGAACCTGG 1201
DB 33515 CGGAAGACAGATTAGAGGAAGCAAAATCCACTGTGTCCTTGGCTGCGGGAACCTGG 33574
QY 1202 ACTAGAGAGTCGCGGCGAGCCCGGAGCCAGCGCTTCGCGCGCTCTAGGCGCGGG 1261
DB 33575 ACTAGAGAGTCGCGGCGAGCCCGGAGCCAGCGCTTCGCGCGCTCTAGGCGCGGG 33634
QY 1262 GCCCGGGGGGAAGGGAGCGAGACCGGGACCCCTAAGACACCTGCTGTACCTCCAC 1321
DB 33635 GCCCGGGGGGAAGGGAGCGAGACCGGGACCCCTAAGACACCTGCTGTACCTCCAC 33694
QY 1322 -----CCCCACCCACACCTCCCTCCCACTCCCTAGATGTGCTGGGGGGCTGAACG 1376
DB 33695 CCCCACCCACACCTCCCTCCCACTCCCTAGATGTGCTGGGGGGCTGAACG 33754
QY 1377 TCGCCCGTTTAAAGGGGGGGGGCGGCTCCAGCTGCTTCTGCTGAGTGAAGTACA 1436
DB 33755 TCGCCCGTTTAAAGGGGGGGGGCGGCTCCAGCTGCTTCTGCTGAGTGAAGTACA 33814
QY 1437 TAAACAGAGCGCGGAACCGGGCGGGAGGAGAGACAGAGCTTTGACCGTAGTAA 1496
DB 33815 TAAACAGAGCGCGGAAGGGGGGGGAGGAGAGACAGAGCTTTGACCGTAGTAA 33874
QY 1497 CCTCTGCTCGGTGTCACCCCGAATCTATPAAAGGAACCTAGTCCCGGCAAAACCCCGTAA 1556
DB 33875 CCTCTGCTCGGTGTCACCCCGAATCTAATAAGGAACCTAGTCTCGGCAAAACCCCGTAA 33934
QY 1557 TTGCAGAGCGAGTGAAGTGGGCGGGAGCCCGAGAGCGGAGCCGCTTCCTCCCGG 1616
DB 33935 TTGCAGAGCGAGTGAAGTGGGCGGGAGCCCGAGAGCGGAGCCGCTTCCTCCCGG 33994
QY 1617 GTGCGGCGAGGCGAGGCGGGAGGAGCTC 1643
DB 33995 GTGCGGCGAGGCGAGGCGGGAGGAGCTC 34021

RESULT 9
AC012230 175064 bp DNA linear HTG 22-APR-2000
LOCUS Homo sapiens clone RP11-1M10, WORKING DRAFT SEQUENCE, 39 unordered
DEFINITION pieces
ACCESSION AC012230
VERSION AC012230.3 GI:7637254
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 175064)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-1M10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175064)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhaltier,B.,
Brown,A., Castle,A., Collangelo,M., Collins,S., Collymore,A.,
Cooper,P., DeArelano,K., Dewar,K., Domingo,M., Donelan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kahn,J., Karatas,A., Klein,J.,
Lehoczky,J., Liou,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLoughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

```

TITLE JOURNAL COMMENT

Petersen,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramaniam,A., Talamas,J.,
Tesfaye,S., Turrel,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 22, 2000 this sequence version replaced gi:6454033.
All repeats were identified using RepeatMasker.

Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: http://www-seg.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L2510
Center clone name: L_M_10

----- Summary Statistics
Sequencing vector: M13: M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731

Consensus quality: 117571 bases at least Q40
Consensus quality: 145749 bases at least Q30
Consensus quality: 160940 bases at least Q20

Insert size: 185000; agarose-fp
Insert size: 171264; sum-of-contigs
Quality coverage: 2.9 in Q20 bases; agarose-fp
Quality coverage: 3.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1003: contig of 1003 bp in length
1004 1103: gap of 100 bp
1104 2634: contig of 1531 bp in length
2635 2734: gap of 100 bp
2735 4415: contig of 1681 bp in length
4416 4515: gap of 100 bp
4516 5785: contig of 1270 bp in length
5786 5885: gap of 100 bp
5886 7879: contig of 1994 bp in length
7880 7979: gap of 100 bp
7980 9686: contig of 1707 bp in length
9687 9786: gap of 100 bp
9787 12253: contig of 2467 bp in length
12254 12353: gap of 100 bp
12354 15228: contig of 2875 bp in length
15229 15328: gap of 100 bp
15329 17200: contig of 1872 bp in length
17201 17300: gap of 100 bp
17301 20131: contig of 2831 bp in length
20132 20231: gap of 100 bp
20232 22587: contig of 2356 bp in length
22588 22687: gap of 100 bp
22688 25707: contig of 3020 bp in length
25708 25807: gap of 100 bp
25808 28184: contig of 2377 bp in length
28185 28284: gap of 100 bp
28285 31338: contig of 3054 bp in length
31339 31438: gap of 100 bp
31439 34299: contig of 2861 bp in length
34300 34399: gap of 100 bp
34400 38318: contig of 3919 bp in length
38319 38418: gap of 100 bp
38419 42835: contig of 4417 bp in length
42836 42935: gap of 100 bp

[illegible]

DeArellano, K., Dewar, K., Domino, M., Doyle, K., Fenestor, J.,
 Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
 Garayna, S., Grant, C., Hagos, B., Heeford, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Landers, T., LePoczky, J., Levine, R., Liu, C., Liu, G., Locke, K.,
 Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
 McPheters, R., Meldrum, J., Menes, L., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connor, J., O'Donnell, P., Oliver, T. M., Peterson, K.,
 Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
 Tirell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
 Zimmer, A., and Zody, N.

TITLE

JOURNAL

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6705871.

COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Greer, P. (1996-1997)

<http://ftp.genome.washington.edu/JM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W13R

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2512

Center clone name: L_N_13

* NOTE: This record contains 73 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 871: contig of 871 bp in length

872 971: gap of 100 bp

972 1834: contig of 863 bp in length

1835 1934: gap of 100 bp

1935 2804: contig of 870 bp in length

2805 2904: gap of 100 bp

2905 3745: contig of 841 bp in length

3746 3845: gap of 100 bp

3846 4696: contig of 851 bp in length

4697 4796: gap of 100 bp

4797 5640: contig of 844 bp in length

5641 5740: gap of 100 bp

5741 6540: contig of 800 bp in length

6541 6640: gap of 100 bp

6641 7503: contig of 869 bp in length

7510 7609: gap of 100 bp

7610 8479: contig of 870 bp in length

8480 8579: gap of 100 bp

8580 9430: contig of 851 bp in length

9431 9530: gap of 100 bp

9531 10376: contig of 846 bp in length

10377 10476: gap of 100 bp

10477 11322: contig of 846 bp in length

11323 11422: gap of 100 bp

11423 12302: contig of 880 bp in length

12303 12402: gap of 100 bp

12403 13280: contig of 878 bp in length

13281 13390: gap of 100 bp

13381 14241: contig of 861 bp in length

14242 14341: gap of 100 bp

14342 15196: contig of 855 bp in length

15197 15296: gap of 100 bp

15297 16123: contig of 827 bp in length

16124 16223: gap of 100 bp

16224 17072: contig of 849 bp in length

17073 17172: gap of 100 bp

17173 18041: contig of 869 bp in length

18042 18141: gap of 100 bp

18142 19009: contig of 868 bp in length

19010 19109: gap of 100 bp

19110 19966: contig of 857 bp in length

19967 20066: gap of 100 bp

20067 20921: contig of 855 bp in length

20922 21021: gap of 100 bp

21022 21865: contig of 844 bp in length

21866 21965: gap of 100 bp

21966 22832: contig of 867 bp in length

22833 22932: gap of 100 bp

22933 23780: contig of 848 bp in length

23781 23880: gap of 100 bp

23881 24733: contig of 853 bp in length

24734 24833: gap of 100 bp

24834 25670: contig of 837 bp in length

25671 25770: gap of 100 bp

25771 26621: contig of 851 bp in length

26622 26721: gap of 100 bp

26722 27576: contig of 855 bp in length

27577 27676: gap of 100 bp

27677 28332: contig of 856 bp in length

28333 28632: gap of 100 bp

28633 29492: contig of 860 bp in length

29493 29592: gap of 100 bp

29593 30455: contig of 863 bp in length

30456 30555: gap of 100 bp

30556 31410: contig of 855 bp in length

31411 31510: gap of 100 bp

31511 32368: contig of 858 bp in length

32369 32468: gap of 100 bp

32469 33312: contig of 844 bp in length

33313 33412: gap of 100 bp

33413 34268: contig of 856 bp in length

34269 34368: gap of 100 bp

34369 35204: contig of 836 bp in length

35205 35304: gap of 100 bp

35305 36156: contig of 852 bp in length

36157 36256: gap of 100 bp

36257 37128: contig of 872 bp in length

37129 37228: gap of 100 bp

37229 38083: contig of 855 bp in length

38084 38183: gap of 100 bp

38184 39031: contig of 848 bp in length

39032 39131: gap of 100 bp

39132 40006: contig of 875 bp in length

40007 40106: gap of 100 bp

40107 40967: contig of 851 bp in length

40968 41067: gap of 100 bp

41068 41913: contig of 846 bp in length

41914 42013: gap of 100 bp

42014 42824: contig of 811 bp in length

42825 42924: gap of 100 bp

42925 43775: contig of 852 bp in length

43776 43876: gap of 100 bp

43877 44752: contig of 876 bp in length

44753 44852: gap of 100 bp

44853 45724: contig of 872 bp in length

45725 45824: gap of 100 bp

45825 46643: contig of 819 bp in length

46644 46743: gap of 100 bp

46744 47599: contig of 856 bp in length

47600 47699: gap of 100 bp

47700 48551: contig of 852 bp in length

48552 48651: gap of 100 bp

48652 49485: contig of 834 bp in length

49486 49585: gap of 100 bp

49586 50440: contig of 855 bp in length

50441 50540: gap of 100 bp

	* * *	50541	5-404:	contig of 864 bp	in length				
	*	5405	51504:	gap of	100 bp				
	*	51505	5372:	contig of 88 bp	in length				
	*	52373	52472:	gap of	100 bp				
	*	52473	53328:	contig of 856 bp	in length				
	*	5329	53428:	gap of	100 bp				
	*	53429	54268:	contig of 840 bp	in length				
	*	54269	54368:	gap of	100 bp				
	*	54369	55229:	contig of 861 bp	in length				
	*	55230	55329:	gap of	100 bp				
	*	55330	56197:	contig of 868 bp	in length				
	*	56198	56297:	gap of	100 bp				
	*	56298	57163:	contig of 866 bp	in length				
	*	57164	57263:	gap of	100 bp				
	*	57264	58130:	contig of 867 bp	in length				
	*	58131	58230:	gap of	100 bp				
	*	58231	59082:	contig of 852 bp	in length				
	*	59083	59182:	gap of	100 bp				
	*	59183	60020:	contig of 838 bp	in length				
	*	60021	60120:	gap of	100 bp				
	*	60211	60983:	contig of 863 bp	in length				
	*	60284	61083:	gap of	100 bp				
	*	61084	61935:	contig of 852 bp	in length				
	*	61936	62035:	gap of	100 bp				
	*	62036	62866:	contig of 831 bp	in length				
	*	62867	62966:	gap of	100 bp				
	*	63267	63827:	contig of 861 bp	in length				
	*	63268	63927:	gap of	100 bp				
	*	63928	64783:	contig of 856 bp	in length				
	*	64784	64883:	gap of	100 bp				
	*	64884	65740:	contig of 857 bp	in length				
	*	65741	65840:	gap of	100 bp				
	*	65841	66684:	contig of 844 bp	in length				
	*	66685	66784:	cap of	100 bp				
	*	66785	67653:	contig of 867 bp	in length				
	*	67652	67751:	gap of	100 bp				
Query Match									
				Best Local Similarity	99.3%; Score 519.8; DB 2;	Length 69570;			
		Matches	543;	Conservative	0;	Mismatches	2;	Gaps	2;
Qy		1099	CCCTCCTTCGGCTGAGAAACTACAAA-GGAATAAAAAATTCGGGAAGCAGCATTTA	1157					
Dd		41109	CCTCTCCTCGCGCTGAGAAACTACAAGGAGAATAAAAAATTCGGGAAGCAGCATTTA	41158					
Qy		1158	GAGGAAGCAAAITCCACTGTGCCCTTGCGTCCCGGGAACGTGGACTAGAGAGTGTGGG	1217					
Dd		41169	GAGGAAGCAAAITCCACTGTGCCCTTGCGTCCCGGGAACGTGGACTAGAGAGTGTGGG	41228					
Qy		1218	CGAGCCCTCGAGCCGACGCCTTCCC CGCGCTTAGCCCGCGCGCCCGCGGGSGGAG	1277					
Dd		41229	CSGAGCCCGAGCCCGACGCCTTCCCGCGCTTAAAGCCGGGGCCCCGGCGGGGERAG	41288					
Qy		1278	GSAGCCGACAGCCGCGACCCTTAAGCACCTGTGTACCTTCCACCCCAACCCCAACCCACC	:337					
Dd		4-289	GGAGCGCAGACCGCGGRACCCCTASACACCTGCTGTACCTCCACCCCAACCCCAACCC	41348					
Qy		1338	TCCGCCCAACT-CCTTAGATGTGTGTGGCGGCTTGAAGTCGCCCGTTTTAAGSGGGGG	1396					
Dd		41349	TCCGCCCAACTGCCCTATATGTGTCTGGCGGCTGAACCTGCCCGTTTTAAGSGGGGG	41408					
Qy		1397	CCCGGGCTCCACGTGCTTTCTGTGTAGTGACTGAACCTGTACCGATAGTAACCTTGC	1516					
Dd		41409	CCCGGGCTCCACGTGCTTTCTGTGTAGTGACTGAACCTGTACCGATAGTAACCTTGC	41528					
Qy		1457	GGCGGGGAGGAGGAGACAGCGCTTTGACCGATAGTAACCTTGCCTCGGTGCGAGCC	1516					
Dd		41469	GGCGGGGAGGAGGAGGAGACAGCGCTTTGACCGATAGTAACCTTGCCTCGGTGCGAGCC	41528					
Qy		1517	GAACTATAAAGAACTAGTCCCGGCAAAACCCCGTAATTTCGAGGGAGAGTCACTGG	1576					
Dd		41539	GAAT						

QY	1577	GGCGGGACCCGAGAGCGAGCCGACCCTTCTCTCCGGGCTTCGCCGCAGGCGAGCGGCGG	1636
Dd	41589	GGCGGGACCCGAGAGCGAGCCGACCCTTCTCTCCGGGCTTCGCCGCAGGCGAGCGGCGG	41648
QY	1637	GGAGCTC	1643
Dd	41649	GGAGCTC	41655
RESULT	13		
LOCUS	AC021246/c		
DEFINITION	Homo sapiens clone RP11-IN10, LOW-PASS SEQUENCE SAMPLING.		
ACCESSION	AC021246		
VERSION	AC021246.2 GI-9113882		
KEYWORDS	HTG; HTGS-PHASE2.		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 69570)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
TITLE	Homo sapiens chromosome, clone RP11-IN10		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 69570)		
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Becker,T., Beka,F., Bogdanovskiy,I., Bonkhgalter,B., Brown,A., Burkett,G., Castle,A., Bozdaylavsky,L., Boukhgalter,B., Collins,S., Collymore,A., Cooke,P., Choepel,Y., Colangelo,M., Collins,S., Doyle,M., Fenestor,J., DeArellano,K., Dewar,K., Dominko,T., Forrest,C., Gage,D., Galagan,J., Ferreira,P., FitzHugh,W., Fodor,C., Heaford,A., Horton,L., Gowda,S., Grant,G., Hagos,B., Kann,L., Karatas,A., Klein,J., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehoczyk,J., Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrim,G., Menes,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K., Pierre,N., Pisani,C., Pollara,Y., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,Z., Tirrell,A., Vassiliev,H., Viel,R., Vo.A., Wu,X., Wymat,D., Ye,W.J., Zimmer,A. and Zody,M.		
TITLE	Direct Submission:		
JOURNAL	Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
COMMENT	On Jul 13, 2000 this sequence version replaced gi:6705871. All repeats were identified using RepeatMasker: Sait,A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html ----- Genome Center Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIBR Web site: http://www-seq.wi.mit.edu Contact: sequence_submissions@genome.wi.mit.edu ----- Project Information Center project name: L2512 Center clone name: L_N10 ----- * NOTE: This record contains 73 individual * sequencing reads that have not been assembled into * contigs. Runs of N are used to separate the reads * and the order in which they appear is completely * arbitrary. Low-pass sequence sampling is useful for * identifying clones that may be gene-rich and allows * overlap relationships among clones to be deduced. * However, it should not be assumed that this clone * will be sequenced to completion. In the event that * the record is updated, the accession number will * be preserved. * 1 971: contig of 871 bp in length * 872 971: gap of 100 bp * 972 1834: contig of 863 bp in length * 1835 934: gap of 100 bp		

```
* 1935 2804: contig of 870 bp in length
* 2805 2904: gap of 100 bp
* 2905 3745: contig of 841 bp in length
* 3746 3845: gap of 100 bp
* 3846 4696: contig of 851 bp in length
* 4697 4796: gap of 100 bp
* 4797 5640: contig of 844 bp in length
* 5641 5740: gap of 100 bp
* 5741 6540: contig of 800 bp in length
* 6541 6640: gap of 100 bp
* 6641 7509: contig of 869 bp in length
* 7510 7609: gap of 100 bp
* 7610 8479: contig of 870 bp in length
* 8480 8579: gap of 100 bp
* 8580 9430: contig of 851 bp in length
* 9431 9530: gap of 100 bp
* 9531 10376: contig of 846 bp in length
* 10377 10476: gap of 100 bp
* 10477 11322: contig of 846 bp in length
* 11323 11422: gap of 100 bp
* 11423 12302: contig of 880 bp in length
* 12303 12402: gap of 100 bp
* 12403 13280: contig of 878 bp in length
* 13281 13380: gap of 100 bp
* 13381 14241: contig of 861 bp in length
* 14242 14341: gap of 100 bp
* 14342 15196: contig of 855 bp in length
* 15197 15296: gap of 100 bp
* 15297 16123: contig of 827 bp in length
* 16124 16223: gap of 100 bp
* 16224 17072: contig of 849 bp in length
* 17073 17172: gap of 100 bp
* 17173 18041: contig of 869 bp in length
* 18042 18141: gap of 100 bp
* 18142 19009: contig of 888 bp in length
* 19010 19109: gap of 100 bp
* 19110 19966: contig of 857 bp in length
* 19967 20066: gap of 100 bp
* 20067 20921: contig of 855 bp in length
* 20922 21021: gap of 100 bp
* 21022 21865: contig of 844 bp in length
* 21866 21965: gap of 100 bp
* 21966 22832: contig of 867 bp in length
* 22833 22932: gap of 100 bp
* 22933 23780: contig of 848 bp in length
* 23781 23880: gap of 100 bp
* 23881 24733: contig of 853 bp in length
* 24734 24833: gap of 100 bp
* 24834 25670: contig of 837 bp in length
* 25671 25770: gap of 100 bp
* 25771 26621: contig of 851 bp in length
* 26622 26721: gap of 100 bp
* 26722 27576: contig of 855 bp in length
* 27577 27676: gap of 100 bp
* 27677 28532: contig of 856 bp in length
* 28533 28632: gap of 100 bp
* 28633 29492: contig of 860 bp in length
* 29493 29592: gap of 100 bp
* 29593 30455: contig of 863 bp in length
* 30456 30555: gap of 100 bp
* 30556 31410: contig of 855 bp in length
* 31411 31510: gap of 100 bp
* 31511 32368: contig of 858 bp in length
* 32369 32468: gap of 100 bp
* 32469 33312: contig of 844 bp in length
* 33313 33412: gap of 100 bp
* 33413 34268: contig of 856 bp in length
* 34269 34368: gap of 100 bp
* 34369 35204: contig of 836 bp in length
* 35205 35304: gap of 100 bp
* 35305 36156: contig of 852 bp in length
* 36157 36256: gap of 100 bp
* 36257 37128: contig of 872 bp in length
```

```
* 37129 37228: gap of 100 bp
* 37229 38083: contig of 855 bp in length
* 38084 38183: gap of 100 bp
* 38184 39031: contig of 848 bp in length
* 39032 39131: gap of 100 bp
* 39132 40006: contig of 875 bp in length
* 40007 40106: gap of 100 bp
* 40107 40967: contig of 861 bp in length
* 40968 41067: gap of 100 bp
* 41069 41913: contig of 846 bp in length
* 41914 42013: gap of 100 bp
* 42014 42824: contig of 811 bp in length
* 42825 42924: gap of 100 bp
* 42925 43776: contig of 852 bp in length
* 43777 43876: gap of 100 bp
* 43877 44752: contig of 876 bp in length
* 44753 44852: gap of 100 bp
* 44853 45724: contig of 872 bp in length
* 45725 45824: gap of 100 bp
* 45825 46643: contig of 819 bp in length
* 46644 46743: gap of 100 bp
* 46744 47599: contig of 856 bp in length
* 47600 47699: gap of 100 bp
* 47700 48551: contig of 852 bp in length
* 48552 48651: gap of 100 bp
* 48652 49485: contig of 834 bp in length
* 49486 49585: gap of 100 bp
* 49586 50440: contig of 855 bp in length
* 50441 50540: gap of 100 bp
* 50541 51404: contig of 864 bp in length
* 51405 51504: gap of 100 bp
* 51505 52372: contig of 868 bp in length
* 52373 52472: gap of 100 bp
* 52473 53328: contig of 856 bp in length
* 53329 53428: gap of 100 bp
* 53429 54268: contig of 840 bp in length
* 54269 54368: gap of 100 bp
* 54369 55229: contig of 861 bp in length
* 55230 55329: gap of 100 bp
* 55330 56197: contig of 868 bp in length
* 56198 56297: gap of 100 bp
* 56298 57163: contig of 866 bp in length
* 57164 57263: gap of 100 bp
* 57264 58130: contig of 867 bp in length
* 58131 58230: gap of 100 bp
* 58231 59082: contig of 852 bp in length
* 59083 59182: gap of 100 bp
* 59183 60020: contig of 838 bp in length
* 60021 60120: gap of 100 bp
* 60121 60983: contig of 863 bp in length
* 60984 61083: gap of 100 bp
* 61084 61935: contig of 852 bp in length
* 61936 62035: gap of 100 bp
* 62036 62866: contig of 831 bp in length
* 62867 62966: gap of 100 bp
* 62967 63827: contig of 861 bp in length
* 63828 63927: gap of 100 bp
* 63928 64783: contig of 856 bp in length
* 64784 64883: gap of 100 bp
* 64884 65740: contig of 857 bp in length
* 65741 65840: gap of 100 bp
* 65841 66684: contig of 844 bp in length
* 66685 66784: gap of 100 bp
* 66785 67651: contig of 867 bp in length
* 67652 67751: gap of 100 bp
```

Query Match

16.5%; Score 271.6; DB 2; Length 69570;

Best Local Similarity 89.8%; Pred. No. 4.9e-61;

Matches 289; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1322 CCCCACCCACCCACCTCCCGCCCAACTCCCTAGATGTGCGGGGGGTGAACGTGGCC 1381

||||| ||| | ||||| | ||||| ||||| ||||| |||||

Db 67642 CCCCCNNCCCCACCACCTCCCGCCCAANTATAGATGTGCGGGGGGTGAACGTGGCC 67583


```
/map="9q31"
/tissue_type="placenta"
1..>697
/gene="ABCA1"
/note="ABCA1"
396..>697
/gene="ABCA1"
/note="membrane-bound"
/codon_start=1
/product="ATP binding cassette transporter 1"
/protein_id="AA269513.1"
/db_xref="GI:7769708"
/translation="MACHPQLRLLWKNLTRRROTQQLLLLEVAMPLEIFLLLSVRL
STFPYEQHECHFPNAMPFAGLPPVQGLICNANNPCFRYPPTGEAPGVVGNFNKS"
BASE COUNT 152 a 198 c 150 g 156 t 1 others
ORIGIN

Query Match 10.5%; Score 173; DB 9; Length 697;
Best Local Similarity 100.0%; Pred.No.5.6e-35;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1471 AGAGCACAGGCTTTGACCGATAGTACCTCTCGCTCGGTGCGACCGCAATCTATAAAAGG 1530
Db 1 AGAGCACAGGCTTTGACCGATAGTACCTCTCGCTCGGTGCGACCGCAATCTATAAAAGG 60

QY 1531 AACTAGTCCCGGCAGAAACCCCTAATTGGAGCGAGAGTGTAGTGGGGCGGACCCGCA 1590
Db 61 AACTAGTCCCGGCAGAAACCCCTAATTGGAGCGAGAGTGTAGTGGGGCGGACCCGCA 120

QY 1591 GAGCCGAGCGGACCCCTCTCTCCCGGGCTGCGGCAGGCGAGGGCGGAGCTC 1643
Db 121 GAGCCGAGCGGACCCCTCTCTCCCGGGCTGCGGCAGGCGAGGGCGGAGCTC 173
```

Search completed: February 4, 2003, 00:59:00
Job time : 5346.97 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 16:38:05 ; Search time 397.962 Seconds
(without alignments)
10165.006 Million cell updates/sec

Title: US-09-596-141C-3_COPY_1394_1532
Perfect score: 139
Sequence: 1 gggcccggtccacgtgct.....agccgaatctataaaaggaa 139

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 2054540 seqs, 14531402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_on.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vt.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vt.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_other.*
33: em_hgt_mus.*
34: em_hgt_pln.*
35: em_hgt_rod.*
36: em_hgt_mam.*
37: em_hgt_vrt.*
38: em_sy.*
39: em_hgtgo_hum.*
40: em_hgtgo_mus.*
41: em_hgtgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	139	100.0	167	9 HSA252201	AJ252201 Homo sapi
2	139	100.0	1643	6 AX060715	AX060715 Sequence
3	139	100.0	1643	6 AX060894	AX060894 Sequence
4	139	100.0	69570	2 AC021246	AC021246 Homo sapi
5	139	100.0	96717	9 AL359182	AL359182 Human DNA
6	139	100.0	175064	2 AC012230	AC012230 Homo sapi
7	139	100.0	183999	6 AX092589	AX092589 Sequence
8	137.4	98.8	1167	9 AF258623s1	AF258623 Homo sapi
9	137.4	98.8	2893	6 AX351031	AX351031 Sequence
10	137.4	98.8	3231	6 AX351029	AX351029 Sequence
11	137.4	98.8	149034	9 AF275948	AF275948 Homo sapi
12	137.4	98.8	201144	9 AF287262	AF287262 Homo sapi
13	126.2	90.8	69570	2 AC021246	AC021246 Homo sapi
14	113.2	81.4	90698	2 AC021345	AC021345 Homo sapi
15	106.2	76.4	186889	2 AL807243	AL807243 Mus muscu
16	106.2	76.4	278572	10 AF287263	AF287263 Mus muscu
17	104	74.8	145833	2 AC125837	AC125837 Rattus no
18	62	44.5	697	9 AF258627	AF258627 Homo sapi
19	36.4	26.2	38	6 AX092823	AX092823 Sequence
20	34.6	24.9	193159	9 AC006946	AC006946 Homo sapi
21	33.4	24.0	129090	2 AP003931	AP003931 Oryza sat
22	33.4	24.0	133902	2 AC102375	AC102375 Mus muscu
23	33.4	24.0	218091	2 AC121802	AC121802 Mus muscu
24	32.8	23.5	107553	9 HSJ136014	HSJ136014 Human DNA
25	32.6	23.5	235198	10 AL645571	AL645571 Mouse DNA
26	32.4	23.3	138151	2 AC094070	AC094070 Rattus no
27	32.4	23.3	150994	2 AC111413	AC111413 Rattus no
28	32.2	23.2	125066	9 AC074286	AC074286 Homo sapi
29	32.2	23.2	156785	2 AC018541	AC018541 Homo sapi
30	32.2	23.2	170924	2 AC087619	AC087619 Homo sapi
31	32.2	23.2	171813	2 AC022722	AC022722 Homo sapi
32	32.2	23.2	185218	2 AC126911	AC126911 Rattus no
33	32.2	23.2	187994	2 AC126146	AC126146 Rattus no
34	32.2	23.2	192139	9 AC021955	AC021955 Homo sapi
35	32.2	23.2	215936	9 CNS05TCA	AL355076 Human chr
36	32	23.0	187359	2 AC068886	AC068886 Homo sapi
37	32	23.0	188425	2 AC026554	AC026554 Homo sapi
38	32	23.0	224229	9 AC011899	AC011899 Homo sapi
39	31.8	22.9	64523	2 AC100987	AC100987 Mus muscu
40	31.8	22.9	164290	2 AC096454	AC096454 Rattus no
41	31.6	22.7	132596	2 AC013046	AC013046 Drosophil
42	31.6	22.7	169101	3 AC008317	AC008317 Drosophil
43	31.6	22.7	212185	9 AC024651	AC024651 Homo sapi
44	31.6	22.7	329362	3 AE003681	AE003681 Drosophil
45	31.4	22.6	215889	2 AC092196	AC092196 Canis fam

ALIGNMENTS

RESULT 1
HSA252201
LOCUS
DEFINITION Homo sapiens partial ABC-1 gene for ATP-binding cassette transporter-1, 5'UTR and promoter region.
ACCESSION AJ252201
VERSION AJ252201.1 GI:12053757
KEYWORDS ABC-1 gene; ATP-binding cassette transporter-1; promoter.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1167)
AUTHORS Persch-Oczumek,M., Langmann,T., Heimerl,S., Borsukova,H., Kaminski,W.E., Drobnik,W., Honer,C., Schumacher,C. and Schmitz,G.

TITLE The zinc finger protein 202 (ZNF202) is a transcriptional repressor of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene expression and a modulator of cellular lipid efflux

JOURNAL MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

1. 1167

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/map="9q22-3"

/cell_type="leukocyte"

1. 1167

/gene="ABC-1"

4...942

/gene="ABC-1"

/function="cholesterol efflux regulatory protein"

896...900

/gene="ABC-1"

1148...1167

/gene="ABC-1"

/function="cholesterol efflux regulatory protein"

278 a 315 c 327 g 247 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 139; DB 9; Length 1167;

Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCCCGCGCTCCACGTGTTCTGCTGAGTGAACATACATAACAGAGCCGGAA 60

|||||

DB 768 GGCCCGCGCTCCACGTGTTCTGCTGAGTGAACATACATAACAGAGCCGGAA 827

|||||

QY 61 CGGGCGGGGAGGAGGAGACAGAGCTTGACCCATAGTACCTCTGGCTCGGTGCA 120

|||||

DB 828 CGGGCGGGGAGGAGGAGACAGAGCTTGACCCATAGTACCTCTGGCTCGGTGCA 120

|||||

QY 121 GCCGAATCTATAAAGGAA 139

|||||

DB 888 GCCGAATCTATAAAGGAA 906

|||||

RESULT 2

AX060715

LOCUS

Sequence 3 from Patent WO0078972.

AX060715

ACCESSION

AX060715.1

VERSION

KEYWORDS

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Law, R.M., Wade, D., and Garvin, M.

Regulation with binding cassette transporter protein: abcl

Patent: WO 0078972-A 3 28-DEC-2000;

CV THERAPEUTICS, INC. (US)

LOCATION/Qualifiers

1..1643

/organism="Homo sapiens"

/db_xref="taxon:9606"

370 a 413 c 457 g 403 t

BASE COUNT

ORIGIN

Query Match

100.0%; Score 139; DB 6; Length 1643;

Best Local Similarity 100.0%; Pred. No. 1.1e-31;

Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCCCGCGCTCCACGTGTTCTGCTGAGTGAACATACATAACAGAGCCGGAA 60

|||||

DB 1394 GGCCCGCGCTCCACGTGTTCTGCTGAGTGAACATACATAACAGAGCCGGAA 1453

|||||

QY 61 CGGGCGGGGAGGAGGAGACAGAGCTTGACCCATAGTACCTCTGGCTCGGTGCA 120

|||||

DB 1454 CGGGCGGGGAGGAGGAGACAGAGCTTGACCCATAGTACCTCTGGCTCGGTGCA 1513

|||||

QY 121 GCCGAATCTATAAAGGAA 139

|||||

DB 1514 GCCGAATCTATAAAGGAA 1532

|||||

RESULT 3

AX060894

LOCUS

Sequence 3 from Patent WO0078971.

AX060894

ACCESSION

AX060894.1

VERSION

KEYWORDS

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Law, R.M., Wade, D., Gram, J.F. and Garvin, M.

ATP binding cassette transporter protein: abcl

Patent: WO 0078971-A 3 28-DEC-2000;

CV THERAPEUTICS, INC. (US)

LOCATION/Qualifiers

1..1643

/organism="Homo sapiens"

/db_xref="taxon:9606"

370 a 413 c 457 g 403 t

BASE COUNT

ORIGIN

Query Match

100.0%; Score 139; DB 6; Length 1643;

Best Local Similarity 100.0%; Pred. No. 1.1e-31;

Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCCCGCGCTCCACGTGTTCTGCTGAGTGAACATACATAACAGAGCCGGAA 60

|||||

DB 1394 GGCCCGCGCTCCACGTGTTCTGCTGAGTGAACATACATAACAGAGCCGGAA 1453

|||||

QY 61 CGGGCGGGGAGGAGGAGACAGAGCTTGACCCATAGTACCTCTGGCTCGGTGCA 120

|||||

DB 1454 CGGGCGGGGAGGAGGAGACAGAGCTTGACCCATAGTACCTCTGGCTCGGTGCA 1513

|||||

QY 121 GCCGAATCTATAAAGGAA 139

|||||

DB 1514 GCCGAATCTATAAAGGAA 1532

|||||

RESULT 4

AC021246

LOCUS

Homo sapiens clone RP11-IN10, LOW-PASS SEQUENCE SAMPLING.

AC021246

ACCESSION

AC021246.2

VERSION

HTG; HTGS-PHASE0.

KEYWORDS

Homo sapiens.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Birken, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome, clone RP11-IN10

Unpublished

REFERENCE

2 (bases 1 to 69570)

AUTHORS

Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

Anderson, S., Baldwin, J., Baroz, N., Beckerly, R., Beda, F.,
 Boguslavsky, L., Bouhgalter, B., Brown, A., Burkett, G., Castle, A.,
 Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,
 DeArrellanc, K., Dewar, K., Domino, M., Doyle, M., Fenesor, J.,
 Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
 Gardyna, S., Grant, G., Haas, B., Heaford, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kane, L., Karatas, A., Klein, J.,
 Landers, T., Lehoczy, J., Levine, R., Lie, C., Liu, G., Locke, K.,
 Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
 McPheeters, R., Meltdrum, J., Meneus, L., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
 Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
 Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
 Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
 Zimmer, A., and Zody, M.

TITLE JOURNAL

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Jul 13, 2000 this sequence version replaced gi:6705871.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RW/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: W13R

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2512

Center clone name: L_N_10

* NOTE: This record contains 73 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 871: contig of 871 bp in length
 872 971: gap of 100 bp
 972 1834: contig of 863 bp in length
 1835 1934: gap of 100 bp
 1935 2804: contig of 870 bp in length
 2805 2904: gap of 100 bp
 2905 3745: contig of 841 bp in length
 3746 3845: gap of 100 bp
 3846 4656: contig of 851 bp in length
 4657 4796: gap of 100 bp
 4797 5640: contig of 844 bp in length
 5641 5740: gap of 100 bp
 5741 6540: contig of 800 bp in length
 6541 6640: gap of 100 bp
 6641 7509: contig of 869 bp in length
 7510 7609: gap of 100 bp
 7610 8479: contig of 870 bp in length
 8480 8579: gap of 100 bp
 8580 9430: contig of 851 bp in length
 9431 9530: gap of 100 bp
 9531 10376: contig of 846 bp in length
 10377 10476: gap of 100 bp
 10477 11322: contig of 846 bp in length
 11323 11422: gap of 100 bp
 11423 12302: contig of 880 bp in length
 12303 12402: gap of 100 bp
 12403 13280: contig of 878 bp in length
 13281 13380: gap of 100 bp
 13381 14241: contig of 861 bp in length
 14242 14341: gap of 100 bp
 14342 15196: contig of 855 bp in length
 15197 15296: gap of 100 bp
 15297 16123: contig of 827 bp in length
 16124 16223: gap of 100 bp
 16224 17072: contig of 849 bp in length
 17073 17172: gap of 100 bp
 17173 18041: contig of 869 bp in length
 18042 18141: gap of 100 bp
 18142 19009: contig of 868 bp in length
 19010 19109: gap of 100 bp
 19110 19966: contig of 857 bp in length
 19967 20066: gap of 100 bp
 20067 20921: contig of 855 bp in length
 20922 21021: gap of 100 bp
 21022 21865: contig of 844 bp in length
 21866 21965: gap of 100 bp
 21966 22832: contig of 867 bp in length
 22833 22932: gap of 100 bp
 22933 23780: contig of 848 bp in length
 23781 23880: gap of 100 bp
 23881 24733: contig of 853 bp in length
 24734 24833: gap of 100 bp
 24834 25670: contig of 837 bp in length
 25671 25770: gap of 100 bp
 25771 26621: contig of 851 bp in length
 26622 26721: gap of 100 bp
 26722 27576: contig of 855 bp in length
 27577 28532: contig of 856 bp in length
 28533 28632: gap of 100 bp
 28633 29492: contig of 860 bp in length
 29493 29592: gap of 100 bp
 29593 30455: contig of 863 bp in length
 30456 30555: gap of 100 bp
 30556 31410: contig of 855 bp in length
 31411 31510: gap of 100 bp
 31511 32368: contig of 858 bp in length
 32369 32468: gap of 100 bp
 32469 33312: contig of 844 bp in length
 33313 33412: gap of 100 bp
 33413 34268: contig of 856 bp in length
 34269 34368: gap of 100 bp
 34369 35204: contig of 836 bp in length
 35205 35304: gap of 100 bp
 35305 36156: contig of 852 bp in length
 36157 36256: gap of 100 bp
 36257 37128: contig of 872 bp in length
 37129 37228: gap of 100 bp
 37229 38083: contig of 855 bp in length
 38084 38183: gap of 100 bp
 38184 39031: contig of 848 bp in length
 39032 39131: gap of 100 bp
 39132 40006: contig of 875 bp in length
 40007 40106: gap of 100 bp
 40107 40967: contig of 861 bp in length
 40968 41067: gap of 100 bp
 41068 41913: contig of 846 bp in length
 41914 42013: gap of 100 bp
 42014 42824: contig of 811 bp in length
 42825 42924: gap of 100 bp
 42925 43776: contig of 852 bp in length
 43777 43876: gap of 100 bp
 43877 44752: contig of 876 bp in length
 44753 44852: gap of 100 bp
 44853 45724: contig of 872 bp in length
 45725 45824: gap of 100 bp
 45825 46643: contig of 819 bp in length
 46644 46743: gap of 100 bp
 46744 47599: contig of 856 bp in length
 47600 47699: gap of 100 bp
 47700 48551: contig of 852 bp in length
 48552 48651: gap of 100 bp
 48652 49485: contig of 834 bp in length


```

QY 121 GCCGAATCTATAAAGGAA 139
      |||||
DQ 17296 GCCGAATCTATAAAGGAA 17278

RESULT 6
AC012230
LOCUS
DEFINITION Homo sapiens clone RP11-IM10, WORKING DRAFT SEQUENCE, 39 unordered
            175064 bp DNA linear HTG 22-APR-2000
            pieces
AC012230
VERSION AC012230.3 GI:7637254
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          1 (bases 1 to 175064)
REFERENCE Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
          Baidwin, J., Barna, N., Beckerly, R., Boguslavskiy, L., Bouckghalter, B.,
          Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
          Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelac, L., Doyle, M.,
          Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
          Galagan, J., Gardyna, S., Grant, S., Hages, B., Heaford, A., Horton, L.,
          Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
          Lehoczy, J., Lied, C., Locke, K., MacDonald, P., Marquis, N.,
          McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrim, J.,
          Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
          Petersen, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severi, P.,
          Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
          Tesfaye, S., Tirell, A., Vassiliev, E., Vo, A., Wheeler, J., Wu, X.,
          Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
          Direct Submission
          Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
          On Apr 22, 2000 this sequence version replaced gi:6454033.
          All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          ----- Genome Center
          Center: Whitehead Institute/ MIT Center for Genome Research
          Center code: WIBR
          Web site: http://www-seg.wi.mit.edu
          Contact: sequence_submissions@genome.wi.mit.edu
          ----- Project Information
          Center project name: L2510
          Center clone name: L_M10
          ----- Summary Statistics
          Sequencing vector: M13; M7815; 100% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version: C.960731
          Consensus quality: 117571 bases at least Q40
          Consensus quality: 145749 bases at least Q30
          Consensus quality: 160940 bases at least Q20
          Insert size: 185000; agarose-fp
          Insert size: 171264; sum-of-contigs
          Quality coverage: 2.9 in Q20 bases; agarose-fp
          Quality coverage: 3.2 in Q20 bases; sum-of-contigs
          -----
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 39 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *
          * 1
          * 1004 1103: gap of 100 bp
          *
          * 1104 2634: contig of 1531 bp in length
          * 2635 2734: gap of 100 bp
          * 2735 4413: contig of 1681 bp in length
          * 4416 4515: gap of 100 bp
          * 4516 5785: contig of 1270 bp in length
          * 5786 5885: gap of 100 bp
          * 5886 7879: contig of 1994 bp in length
          * 7880 7979: gap of 100 bp
          * 7980 9686: contig of 1707 bp in length
          * 9687 9786: gap of 100 bp
          * 9787 12253: contig of 2467 bp in length
          * 12254 12353: gap of 100 bp
          * 12354 15228: contig of 2875 bp in length
          * 15229 15328: gap of 100 bp
          * 15329 17200: contig of 1872 bp in length
          * 17201 17300: gap of 100 bp
          * 17301 20131: contig of 2831 bp in length
          * 20132 20231: gap of 100 bp
          * 20232 22587: contig of 2356 bp in length
          * 22588 22687: gap of 100 bp
          * 22688 25707: contig of 3020 bp in length
          * 25708 25807: gap of 100 bp
          * 25808 28184: contig of 2377 bp in length
          * 28185 28284: gap of 100 bp
          * 28285 31338: contig of 3054 bp in length
          * 31339 31438: gap of 100 bp
          * 31439 34299: contig of 2861 bp in length
          * 34300 34399: gap of 100 bp
          * 34400 36318: contig of 3919 bp in length
          * 36319 38418: gap of 100 bp
          * 38419 42835: contig of 4417 bp in length
          * 42836 42935: gap of 100 bp
          * 42936 45448: contig of 2513 bp in length
          * 45449 45548: gap of 100 bp
          * 45549 48117: contig of 2568 bp in length
          * 48117 48216: gap of 100 bp
          * 48217 52618: contig of 4402 bp in length
          * 52619 52718: gap of 100 bp
          * 52719 56592: contig of 3874 bp in length
          * 56593 56692: gap of 100 bp
          * 56693 59635: contig of 2943 bp in length
          * 59636 59735: gap of 100 bp
          * 59736 63661: contig of 3926 bp in length
          * 63662 63761: gap of 100 bp
          * 63762 68437: contig of 4676 bp in length
          * 68438 68537: gap of 100 bp
          * 68538 71458: contig of 2921 bp in length
          * 71459 71558: gap of 100 bp
          * 71559 76888: contig of 5330 bp in length
          * 76889 76988: gap of 100 bp
          * 76989 82113: contig of 5125 bp in length
          * 82114 82213: gap of 100 bp
          * 82214 88220: contig of 6007 bp in length
          * 88221 88320: gap of 100 bp
          * 88321 93489: contig of 5179 bp in length
          * 93500 93599: gap of 100 bp
          * 93600 97901: contig of 4302 bp in length
          * 97902 98001: gap of 100 bp
          * 98002 103016: contig of 5015 bp in length
          * 103017 103116: gap of 100 bp
          * 103117 109178: contig of 6062 bp in length
          * 109179 109278: gap of 100 bp
          * 109279 117307: contig of 8029 bp in length
          * 117308 117407: gap of 100 bp
          * 117408 124079: contig of 6672 bp in length
          * 124080 124179: gap of 100 bp
          * 124180 131281: contig of 7102 bp in length
          * 131282 131381: gap of 100 bp
          * 131382 138059: contig of 6678 bp in length
          * 138060 138159: gap of 100 bp
          * 138160 145491: contig of 7332 bp in length
          * 145492 145591: gap of 100 bp
          * 145592 157391: contig of 11800 bp in length

```



```

3 (bases 1 to 1167;
Pallinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C.,
Aouizerat,B.E., Fielding,C.C., and Kane,J.P.
Direct Submission
Submitted (23-JUN-2000) Cardiovascular Research Institute,
University of California, San Francisco, 505 Parnassus Avenue, San
Francisco, CA 94143-0130, USA
Sequence update by submitter
REMARK
COMMENT On Jun 23, 2000 this sequence version replaced gi:7769713.
FEATURES
source
    1..1167
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="9"
        /map="9c31"
        /number=1
    promoter
        224..644
        /gene="ABCA1"
    exon
        845..1147
        /gene="ABCA1"
BASE COUNT 272 a 313 c 328 g 248 t
ORIGIN
Query Match 98.8%; Score 137.4; DB 9; Length 1167;
Best Local Similarity 99.3%; Pred. No. 3.7e-31;
Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GGGCCCGGGTCCAGTGCCTTCTGCTGAGTGACTGAACACTACATAAAGAGCGGGGAA 60
Db 768 GGGCCCGGGTCCAGTGCCTTCTGCTGAGTGACTGAACACTACATAAAGAGCGGGGAA 827
QY 61 CGGGCCGGGAGGAGGAGACAGAGCGGCTTTGACCGATAGTAACCTCTGCGCTCGGIGCA 120
Db 828 GGGCCGGGAGGAGGAGACAGAGCGGCTTTGACCGATAGTAACCTCTGCGCTCGGIGCA 887
QY 121 GCCGAATCTATAAAGGAA 139
Db 888 GCCGAATCTATAAAGGAA 906
RESULT 9
AX351031
LOCUS AX351031 2893 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 3 from Patent WO0183746.
ACCESSION AX351031
VERSION AX351031.1 GI:18616387
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Rosier-Montus,M.F., Prades,C., Lemoine,C., Naudin,L., Deneffe,P.,
Brewer,B., Duverger,N., Remaley,A. and Santamarina-Fojo,S.
Regulatory nucleic acid sequences of the abcl gene
JOURNAL Patent: WO 0183746-A 3 08-NOV-2001;
Aventis Pharma S.A. (FR)
FEATURES
source
    1..2893
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
BASE COUNT 745 a 681 c 769 g 698 t
ORIGIN
Query Match 98.8%; Score 137.4; DB 6; Length 2893;
Best Local Similarity 99.3%; Pred. No. 3.3e-31;
Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GGGCCCGGGTCCAGTGCCTTCTGCTGAGTGACTGAACACTACATAAAGAGCGGGGAA 60
Db 2735 GGGCCCGGGTCCAGTGCCTTCTGCTGAGTGACTGAACACTACATAAAGAGCGGGGAA 2794
QY 61 CGGGCCGGGAGGAGACAGAGCGGCTTTGACCGATAGTAACCTCTGCGCTCGGTGCA 120

```

```

Db 2795 GGGCCGGGAGGAGGAGACAGAGCGGCTTTGACCGATAGTAACCTCTGCGCTCGGTGCA 2854
QY 121 GCCGAATCTATAAAGGAA 139
Db 2855 GCCGAATCTATAAAGGAA 2873
RESULT 10
AX351029
LOCUS AX351029 3231 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 1 from Patent WO0183746.
ACCESSION AX351029
VERSION AX351029.1 GI:18616385
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Rosier-Montus,M.F., Prades,C., Lemoine,C., Naudin,L., Deneffe,P.,
Brewer,B., Duverger,N., Remaley,A. and Santamarina-Fojo,S.
Regulatory nucleic acid sequences of the abcl gene
JOURNAL Patent: WO 0183746-A 1 08-NOV-2001;
Aventis Pharma S.A. (FR)
FEATURES
source
    1..3231
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
BASE COUNT 809 a 773 c 876 g 773 t
ORIGIN
Query Match 98.8%; Score 137.4; DB 6; Length 3231;
Best Local Similarity 99.3%; Pred. No. 3.3e-31;
Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GGGCCCGGGTCCAGTGCCTTCTGCTGAGTGACTGAACACTACATAAAGAGCGGGGAA 60
Db 2735 GGGCCCGGGTCCAGTGCCTTCTGCTGAGTGACTGAACACTACATAAAGAGCGGGGAA 2794
QY 61 CGGGCCGGGAGGAGGAGACAGAGCGGCTTTGACCGATAGTAACCTCTGCGCTCGGTGCA 120
Db 2795 GGGCCGGGAGGAGGAGACAGAGCGGCTTTGACCGATAGTAACCTCTGCGCTCGGTGCA 2834
QY 121 GCCGAATCTATAAAGGAA 139
Db 2855 GCCGAATCTATAAAGGAA 2873
RESULT 11
AP275948
LOCUS AP275948 149034 bp DNA linear PRI 17-JUL-2000
DEFINITION Homo sapiens ABCA1 (ABCA1) gene, complete cds.
ACCESSION AP275948
VERSION AP275948.1 GI:9247085
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 149034)
Santamarina-Fojo,S., Peterson,K., Knapper,C., Qiu,Y., Freeman,L.,
Cheng,J.F., Oscio,J., Remaley,A., Yang,X.P., Haudenschild,C.,
Prades,C., Chimini,G., Blackmon,E., Francois,I., Duverger,N.,
Rubin,E.M., Rosier,M., Deneffe,P., Fredrickson,D.S. and Brewer,H.B.
Jr.
Complete genomic sequence of the human ABCA1 gene: analysis of the
human and mouse ATP-binding cassette A promoter
Proc. Natl. Acad. Sci. U.S.A. 97 (14), 7987-7992 (2000)
JOURNAL 20345099
MEDLINE 10884428
FOBMED
REFERENCE
2 (bases 1 to 149034)
Santamarina-Fojo,S., Peterson,K.M., Knapper,C.L., Freeman,L.A.,

```

July 5th

Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lebozky, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Morrow, J., Nay, G., Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, T.M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J., Zimmer, A., and Zody, M.

Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6705671.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L2512

Center clone name: L_N_10

* NOTE: This record contains 73 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 872 971: contig of 871 bp in length
 972 971: gap of 100 bp
 1835 1934: contig of 863 bp in length
 1935 1934: gap of 100 bp
 2805 2904: contig of 870 bp in length
 2905 2904: gap of 100 bp
 3746 3845: contig of 841 bp in length
 3846 3845: gap of 100 bp
 4697 4796: contig of 851 bp in length
 4797 4796: gap of 100 bp
 5641 5740: contig of 844 bp in length
 5741 5740: gap of 100 bp
 6541 6640: contig of 800 bp in length
 6641 6640: gap of 100 bp
 7510 7609: contig of 869 bp in length
 7610 7609: gap of 100 bp
 8480 8579: contig of 870 bp in length
 8580 8579: gap of 100 bp
 9431 9530: contig of 851 bp in length
 9531 9530: gap of 100 bp
 10377 10476: contig of 846 bp in length
 10477 10476: gap of 100 bp
 11327 11322: contig of 846 bp in length
 11323 11422: gap of 100 bp
 11423 12302: contig of 880 bp in length
 12303 12402: gap of 100 bp
 12403 13280: contig of 878 bp in length
 13281 13380: gap of 100 bp
 13381 14241: contig of 861 bp in length
 14242 14341: gap of 100 bp
 14342 15196: contig of 855 bp in length
 15197 15295: gap of 100 bp
 15297 16123: contig of 827 bp in length
 16124 16223: gap of 100 bp
 16224 17072: contig of 849 bp in length

* 17073 17172: gap of 100 bp
 * 17173 18041: contig of 869 bp in length
 * 18042 18141: gap of 100 bp
 * 18142 19003: contig of 868 bp in length
 * 19010 19109: gap of 100 bp
 * 19110 19966: contig of 857 bp in length
 * 19967 20066: gap of 100 bp
 * 20067 20921: contig of 855 bp in length
 * 20922 21021: gap of 100 bp
 * 21022 21865: contig of 844 bp in length
 * 21866 21965: gap of 100 bp
 * 21966 22832: contig of 867 bp in length
 * 22833 22932: gap of 100 bp
 * 22933 23780: contig of 848 bp in length
 * 23781 23880: gap of 100 bp
 * 23881 24733: contig of 853 bp in length
 * 24734 24833: gap of 100 bp
 * 24834 25670: contig of 837 bp in length
 * 25671 25770: gap of 100 bp
 * 25771 26621: contig of 851 bp in length
 * 26622 26721: gap of 100 bp
 * 26722 27576: contig of 855 bp in length
 * 27577 27676: gap of 100 bp
 * 27677 28532: contig of 856 bp in length
 * 28533 28632: gap of 100 bp
 * 28633 29492: contig of 860 bp in length
 * 29493 29592: gap of 100 bp
 * 29593 30455: contig of 863 bp in length
 * 30456 30555: gap of 100 bp
 * 30556 31410: contig of 855 bp in length
 * 31411 31510: gap of 100 bp
 * 31511 32358: contig of 858 bp in length
 * 32359 32468: gap of 100 bp
 * 32469 33312: contig of 844 bp in length
 * 33313 33412: gap of 100 bp
 * 33413 34288: contig of 856 bp in length
 * 34269 34368: gap of 100 bp
 * 34369 35204: contig of 836 bp in length
 * 35205 35304: gap of 100 bp
 * 35305 36156: contig of 852 bp in length
 * 36157 36256: gap of 100 bp
 * 36257 37128: contig of 872 bp in length
 * 37129 37228: gap of 100 bp
 * 37229 38083: contig of 855 bp in length
 * 38084 38183: gap of 100 bp
 * 38184 39031: contig of 848 bp in length
 * 39032 39131: gap of 100 bp
 * 39132 40006: contig of 875 bp in length
 * 40007 40106: gap of 100 bp
 * 40107 40967: contig of 861 bp in length
 * 40968 41067: gap of 100 bp
 * 41068 41913: contig of 846 bp in length
 * 41914 42013: gap of 100 bp
 * 42014 42824: contig of 811 bp in length
 * 42825 42924: gap of 100 bp
 * 42925 43776: contig of 852 bp in length
 * 43777 43876: gap of 100 bp
 * 43877 44752: contig of 876 bp in length
 * 44753 44852: gap of 100 bp
 * 44853 45724: contig of 872 bp in length
 * 45725 45824: gap of 100 bp
 * 45825 46643: contig of 815 bp in length
 * 46644 46743: gap of 100 bp
 * 46744 47599: contig of 856 bp in length
 * 47600 47699: gap of 100 bp
 * 47700 48551: contig of 852 bp in length
 * 48552 48651: gap of 100 bp
 * 48652 49485: contig of 834 bp in length
 * 49486 49585: gap of 100 bp
 * 49586 50440: contig of 855 bp in length
 * 50441 50540: gap of 100 bp
 * 50541 51404: contig of 864 bp in length
 * 51405 51504: gap of 100 bp

```

* 51505 52372: contig of 968 bp in length
* 52373 52472: gap of 100 bp
* 52473 53238: contig of 856 bp in length
* 53329 53428: gap of 100 bp
* 53429 54268: contig of 840 bp in length
* 54269 54368: gap of 100 bp
* 54369 55229: contig of 861 bp in length
* 55230 55329: gap of 100 bp
* 55330 56197: contig of 868 bp in length
* 56198 56297: gap of 100 bp
* 56298 57163: contig of 866 bp in length
* 57164 57263: gap of 100 bp
* 57264 58130: contig of 867 bp in length
* 58131 58230: gap of 100 bp
* 58231 59082: contig of 852 bp in length
* 59083 59182: gap of 100 bp
* 59183 60020: contig of 838 bp in length
* 60021 60120: gap of 100 bp
* 60121 60983: contig of 863 bp in length
* 60984 61083: gap of 100 bp
* 61084 61935: contig of 852 bp in length
* 61936 62035: gap of 100 bp
* 62036 62866: contig of 831 bp in length
* 62867 62966: gap of 100 bp
* 62967 63827: contig of 861 bp in length
* 63828 63927: gap of 100 bp
* 63928 64783: contig of 856 bp in length
* 64784 64883: gap of 100 bp
* 64884 65740: contig of 857 bp in length
* 65741 65840: gap of 100 bp
* 65841 66684: contig of 844 bp in length
* 66685 66784: gap of 100 bp
* 66785 67651: contig of 867 bp in length
* 67652 67751: gap of 100 bp

Query Match: 90.88; Score 126.2; DB 2; Length 69570;
Best Local Similarity 94.23; Pred. No. 5.8e-28;
Matches 131; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 GGGCCCGGCTCCAGTCTTCTGCTGAGTGACTGAACATACATAACAGCGCGGAA 60
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 67570 GGGCCCGGCTACAGTCTTGTGCTGAGTGACTGAACATACATACAGCGCGGAA 67511

QY 61 CGGGGGGGGAGGAGGAGACAGCGGTTTGACCGATAGTAACCTCGCTCGGTSCA 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 67510 CGGGGGGGGAGGAGGAGACAGCGGTTTGACTGTAGTAGTAACCTCGCTCGGTSCA 67451

QY 121 GCCGAATCTATAAAGGAA 139
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 67450 GCCGAATCTATAAAGGAA 67432

RESULT 14
AC021345/c
LOCUS AC021345 90688 bp DNA linear HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-24J9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC021345
VERSION AC021345.2 GI:9130845
KEYWORDS HTG; HTGS-PHASED.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 90698)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Aldrich,K.,
TITLE Birren,B., Linton,L., Nusbaum,C., Lander,E., Aldrich,K.,
JOURNAL Anderson,S., Baldwin,J., Barna,N., Becker,Y., Bieda,F.,
REFERENCE Boguslavsky,I., Boukhaltier,B., Brown,A., Burkett,G., Castle,A.,
AUTHORS Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,

```

Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Landers,T., Lehoczy,J., Levine,C., Lieu,C., Liu,G., Locke,K.,
 Macdonald,P., Marguis,N., McEwan,P., McGurk,A., McKernan,K.,
 McPheeters,R., Meldrum,J., Menes,L., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., C'Donnell,P., Olivari,T.M., Peterson,K.,
 Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
 Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Tjrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 13, 2000 this sequence version replaced gi:6705761.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WISR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L4483
 Center clone name: 24_J_9

 * NOTE: this record contains 92 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved. 910: contig of 910 bp in length
 * 911 1010: gap of 100 bp
 * 1011 1873: contig of 863 bp in length
 * 1874 1973: gap of 100 bp
 * 1974 2824: contig of 851 bp in length
 * 2825 2924: gap of 100 bp
 * 2925 3802: contig of 878 bp in length
 * 3803 3902: gap of 100 bp
 * 3903 4816: contig of 914 bp in length
 * 4817 4916: gap of 100 bp
 * 4917 5759: contig of 843 bp in length
 * 5760 5859: gap of 100 bp
 * 5860 6764: contig of 905 bp in length
 * 6765 6864: gap of 100 bp
 * 6865 7747: contig of 883 bp in length
 * 7748 7847: gap of 100 bp
 * 7848 8755: contig of 908 bp in length
 * 8756 8855: gap of 100 bp
 * 8856 9753: contig of 898 bp in length
 * 9754 9853: gap of 100 bp
 * 9854 10757: contig of 904 bp in length
 * 10758 10857: gap of 100 bp
 * 10858 11732: contig of 875 bp in length
 * 11733 11832: gap of 100 bp
 * 11833 12739: contig of 907 bp in length
 * 12740 12839: gap of 100 bp
 * 12840 13710: contig of 871 bp in length
 * 13711 13810: gap of 100 bp
 * 13811 14684: contig of 874 bp in length
 * 14685 14784: gap of 100 bp
 * 14785 15602: contig of 878 bp in length
 * 15603 15762: gap of 100 bp
 * 15763 16677: contig of 915 bp in length
 * 16678 16777: gap of 100 bp

```

* 16778 17678: contig of 901 bp in length
* 17679 17778: gap of 100 bp
* 1779 18679: contig of 901 bp in length
* 18680 18779: gap of 100 bp
* 1880 19632: contig of 853 bp in length
* 19633 19732: gap of 100 bp
* 19733 20634: contig of 902 bp in length
* 20635 20734: gap of 100 bp
* 20735 21620: contig of 886 bp in length
* 21621 21720: gap of 100 bp
* 21721 22579: contig of 859 bp in length
* 22580 22679: gap of 100 bp
* 22680 23568: contig of 889 bp in length
* 23569 23668: gap of 100 bp
* 23669 24554: contig of 886 bp in length
* 24555 24654: gap of 100 bp
* 24655 25521: contig of 867 bp in length
* 25522 25621: gap of 100 bp
* 25622 26487: contig of 866 bp in length
* 26488 26587: gap of 100 bp
* 26588 27464: contig of 877 bp in length
* 27465 27564: gap of 100 bp
* 27565 28466: contig of 902 bp in length
* 28467 28566: gap of 100 bp
* 28567 29464: contig of 896 bp in length
* 29465 29564: gap of 100 bp
* 29565 30447: contig of 883 bp in length
* 30448 30547: gap of 100 bp
* 30548 31453: contig of 906 bp in length
* 31454 31553: gap of 100 bp
* 31554 32452: contig of 899 bp in length
* 32453 32552: gap of 100 bp
* 32553 33447: contig of 895 bp in length
* 33448 33547: gap of 100 bp
* 33548 34435: contig of 888 bp in length
* 34436 34535: gap of 100 bp
* 34536 35433: contig of 898 bp in length
* 35434 35533: gap of 100 bp
* 35534 36440: contig of 907 bp in length
* 36441 36540: gap of 100 bp
* 36541 37422: contig of 882 bp in length
* 37423 37522: gap of 100 bp
* 37523 38402: contig of 880 bp in length
* 38403 38502: gap of 100 bp
* 38503 39380: contig of 878 bp in length
* 39381 39480: gap of 100 bp
* 39481 40407: contig of 927 bp in length
* 40408 40507: gap of 100 bp
* 40508 41405: contig of 898 bp in length
* 41406 41505: gap of 100 bp
* 41506 42399: contig of 894 bp in length
* 42400 42499: gap of 100 bp
* 42500 43357: contig of 858 bp in length
* 43358 43457: gap of 100 bp
* 43458 44356: contig of 899 bp in length
* 44357 44456: gap of 100 bp
* 44457 45325: contig of 869 bp in length
* 45326 45425: gap of 100 bp
* 45426 46305: contig of 880 bp in length
* 46306 46405: gap of 100 bp
* 46406 47302: contig of 897 bp in length
* 47303 47402: gap of 100 bp
* 47403 48293: contig of 891 bp in length
* 48294 48393: gap of 100 bp
* 48394 49280: contig of 887 bp in length
* 49281 49380: gap of 100 bp
* 49381 50250: contig of 870 bp in length
* 50251 50350: gap of 100 bp
* 50351 51209: contig of 859 bp in length
* 51210 51309: gap of 100 bp
* 51310 52221: contig of 912 bp in length
* 52222 52321: gap of 100 bp
* 52322 53205: contig of 884 bp in length

```

```

* 53206 53305: gap of 100 bp
* 53306 54225: contig of 920 bp in length
* 54226 54325: gap of 100 bp
* 54326 55206: contig of 881 bp in length
* 55207 55306: gap of 100 bp
* 55307 56165: contig of 859 bp in length
* 56166 56265: gap of 100 bp
* 56266 57124: contig of 859 bp in length
* 57125 57224: gap of 100 bp
* 57225 58121: contig of 897 bp in length
* 58122 58221: gap of 100 bp
* 58222 59075: contig of 854 bp in length
* 59076 59175: gap of 100 bp
* 59176 60058: contig of 883 bp in length
* 60059 60158: gap of 100 bp
* 60159 61067: contig of 909 bp in length
* 61068 61167: gap of 100 bp
* 61168 62051: contig of 884 bp in length
* 62052 62151: gap of 100 bp
* 62152 63022: contig of 871 bp in length
* 63023 63122: gap of 100 bp
* 63123 64026: contig of 904 bp in length
* 64027 64126: gap of 100 bp
* 64127 65015: contig of 889 bp in length
* 65016 65115: gap of 100 bp
* 65116 65975: contig of 860 bp in length
* 65976 66075: gap of 100 bp
* 66076 66954: contig of 879 bp in length
* 66955 67054: gap of 100 bp
* 67055 67947: contig of 893 bp in length
* 67948 68047: gap of 100 bp
* 68048 68934: contig of 887 bp in length
* 68935 69034: gap of 100 bp
* 69035 69910: contig of 876 bp in length
* 69911 70010: gap of 100 bp

```

```

Query Match      81.4%; Score 113.2; DB 2; Length 90698;
Best Local Similarity 92.8%; Pred. No. 5e-24;
Matches 129; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

```

```

QY 1 GGGCCCGCGCTCCACGCTTTCTGCTGAGTGACTGAACTACATAACACAGACGCCGGAA 60
    |||||
Db 87691 GGGCCCGCGCTCCACGCTTTCTGCTGAGTGCTGGAACATACATAACAGA-GCCCGAA 87633
    |||||
QY 61 CGGGCGGGGAGGAGGAGACACAGCTTGACCGTAGTACCTCGCGCTCGGCGCA 120
    |||||
Db 87632 CGGGCGGGGAGGAGGAGACACAGCTTTGACGNATAGTACCTCGCGCTCGGCGCA 87573
    |||||
QY 121 GCCGAATCTATAAAGGAA 139
    |||||
Db 87572 GCCGAATCTATAAAGGAA 87554
    |||||

```

RESULT 15

```

AL807243/c
LOCUS      186889 bp DNA linear HTG 30-JUN-2002
DEFINITION Mus musculus chromosome 4 clone RP23-25D17, *** SEQUENCING IN
PROGRESS ***; 24 unordered pieces.
ACCESSION AL807243
VERSION AL807243.5 GI:21668234
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
            McLay, K.
REFERENCE 1 (bases 1 to 186889)
AUTHORS Direct Submission
TITLE Submitted (29-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
        tumquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
        On Jul 2, 2002 this sequence version replaced gi:21668136.
COMMENT ----- Genome Center

```


Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humphries@sanger.ac.uk
 ----- Project Information
 Center project name: BM25D17

----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 175959 bases at least Q40
 Consensus quality: 179463 bases at least Q30
 Consensus quality: 182069 bases at least Q20
 Insert size: 184589; sum-of-contigs
 Insert size: 189466; 5.5% error; agarose-fp
 Quality coverage: 4.72x in Q20 bases; sum-of-contigs Quality
 coverage: 4.82x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 24 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1
4850 4949: gap of 100 bp
4950 22620: contig of 17671 bp in length
22621 22720: gap of 100 bp
22721 29963: contig of 7243 bp in length
29964 30053: gap of 100 bp
30054 40087: contig of 10024 bp in length
40088 40187: gap of 100 bp
40188 44438: contig of 4251 bp in length
44439 44538: gap of 100 bp
44539 48960: contig of 4422 bp in length
48961 49050: gap of 100 bp
49051 56007: contig of 6947 bp in length
56008 56107: gap of 100 bp
56108 58264: contig of 2177 bp in length
58265 58384: gap of 100 bp
58385 60786: contig of 2402 bp in length
60787 60886: gap of 100 bp
60887 66707: contig of 5821 bp in length
66708 66807: gap of 100 bp
66808 92102: contig of 25295 bp in length
92103 92202: gap of 100 bp
92203 99909: contig of 7707 bp in length
99910 100009: gap of 100 bp
100010 103760: contig of 3751 bp in length
103761 103860: gap of 100 bp
103861 110765: contig of 6905 bp in length
110766 110865: gap of 100 bp
110866 114685: contig of 3820 bp in length
114686 114785: gap of 100 bp
114786 125959: contig of 11174 bp in length
125960 126059: gap of 100 bp
126060 126749: contig of 690 bp in length
126750 128849: gap of 100 bp
128850 132373: contig of 3524 bp in length
132374 132473: gap of 100 bp
132474 134930: contig of 2457 bp in length
134931 135030: gap of 100 bp
135031 139301: contig of 4271 bp in length
139302 139401: gap of 100 bp
139402 144448: contig of 5047 bp in length
144449 144548: gap of 100 bp
144549 147074: contig of 2526 bp in length
147075 147174: gap of 100 bp
147175 156317: contig of 9143 bp in length
156318 156417: gap of 100 bp
156418 186889: contig of 30472 bp in length.
Location/Qualifiers

```

FEATURES

```

Source
1..186889
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone_lib="RP23-25D17"
/clone_lib="RP23-25D17"
1..4849
/feature="assembly_fragment:01920"
fragment_chain:1
4950..22620
/feature="assembly_fragment:01761"
fragment_chain:1
22721..29963
/feature="assembly_fragment:00140"
fragment_chain:1
30064..40087
/feature="assembly_fragment:00335"
fragment_chain:1
40188..44438
/feature="assembly_fragment:00241"
fragment_chain:1
44539..48960
/feature="assembly_fragment:01430"
fragment_chain:1
49061..56007
/feature="assembly_fragment:01667"
fragment_chain:1
56108..58284
/feature="assembly_fragment:00965"
fragment_chain:2
58385..60786
/feature="assembly_fragment:01660"
fragment_chain:2
60887..66707
/feature="assembly_fragment:00374"
fragment_chain:2
66808..92102
/feature="assembly_fragment:01792"
fragment_chain:2
92203..99909
/feature="assembly_fragment:01082"
fragment_chain:2
100010..103760
/feature="assembly_fragment:00358"
fragment_chain:2
103861..110765
/feature="assembly_fragment:01182"
fragment_chain:3
110866..114685
/feature="assembly_fragment:01582"
fragment_chain:3
114786..125959
/feature="assembly_fragment:00043"
fragment_chain:3
126060..128749
/feature="assembly_fragment:00352"
fragment_chain:3
128850..132373
/feature="assembly_fragment:00217"
fragment_chain:4
132474..134930
/feature="assembly_fragment:01737"
fragment_chain:4
135031..139301
/feature="assembly_fragment:00487"
139402..144448
/feature="assembly_fragment:00557"
144549..147074
/feature="assembly_fragment:01342"
147175..156317
/feature="assembly_fragment:01564.0"
156418..186889
/feature="assembly_fragment:01773"

```

BASE COUNT 52900 a 39593 c 40710 g 51338 t 2348 others
ORIGIN
Query Match 76.4%; Score 106.2; DB 2; Length 188889;
Best Loca. Similarity 89.4%; Pred. No. 6.1e-22;
Matches 126; Conservative 0; Mismatches 13; Indels 2; Gaps 1;
Qy 1 GGGCCCCGGCTCCAGTGTCTTCTGAGTGACTGAACATACATAAACAGAGCGCGGAA 60
Db 99260 GGGCCAICTCTCCAGTGTCTTCTGAGTGACTGAACATACATAAACAGAGCGCGGAA 99201
Qy 61 CGGGGGGGG--AGGAGGGAGGACAGGCTTTGACCGATAGTACCTCTGGGCTGGTG 118
Db 99200 GGGGGGGGGGAAAGAGGGAGAGACAGCGTTGACCGGTAGTACCCCGCGCTCGGA 99141
Qy 119 CAGCCGAATCTATAAAGGAA 139
Db 99140 CAGCCGAATCTATAAAGGAA 99120

Search completed: February 4, 2003, 01:43:20
Job time : 748.962 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Computer Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 16:39:45 ; Search time 3108.84 Seconds
(without alignments)
8559.210 Million cell updates/sec

Title: US-09-596-141c-3

Perfect score: 1643

Sequence: 1 gatttccttgctggtgctc.....cagggcagggcgggagcgc 1643

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST: *

1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_estl: *
10: gb_est2: *
11: gb_hic: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	245.4	14.9	289	10	AW748338 RC6-BT025
2	111.4	6.8	467	9	AA527406 ng37502.s
3	89.6	5.4	736	9	AU135588 AU135588
4	89.6	5.4	547	9	AL698654 DRF26686N
5	89.6	5.4	763	9	AU121731 AU121731
6	75	4.6	292	14	Z44377 HSC125081.3

7	75	4.6	1004	11	BC034824	Homo sapi
8	60.8	3.7	1201	17	CNS015W3	Drosophila
9	60.2	3.7	998	12	BC678861	Bombus terrestris
10	54.8	3.3	298	17	CNS04M2X	Tetradon
11	53.8	3.3	844	17	CNS0052P	Drosophila
12	53	3.2	237	9	AU222489	AU222489
C 13	52.4	3.2	910	17	CNS006ON	Drosophila
14	51.2	3.1	832	17	CNS015XC	Drosophila
C 15	51	3.1	681	17	CNS02FE9	Tetradon
C 16	50.6	3.1	925	17	CNS0091P	AL053613 Drosophila
C 17	50.6	3.1	1101	17	CNS017S1	AL108460 Drosophila
C 18	50.4	3.1	944	17	CNS0052P	AL056652 Drosophila
C 19	50.4	3.1	970	17	CNS005E2	AL064371 Drosophila
C 20	48.4	2.9	1304	14	BQ944204	BQ944204 AGENCOURT
C 21	48	2.9	900	17	AG081217	AG081217 Pan trogl
C 22	47.5	2.9	1009	17	CNS010E3	AL098882 Drosophila
C 23	47.6	2.9	1101	17	CNS01720	AL107514 Drosophila
C 24	47.4	2.9	1059	14	BQ680479	BQ680479 AGENCOURT
C 25	47.4	2.9	1101	17	CNS017V2	AL108536 Drosophila
C 26	47.2	2.9	477	9	AL513813	AL513813
C 27	47.2	2.9	925	17	CNS0091P	AL053013 Drosophila
C 28	47.2	2.9	942	17	AG042878	AG042878 Pan trogl
C 29	47	2.9	658	12	BF968907	BF968907 602270874
C 30	47	2.9	843	17	CNS00CS1	AL059666 Drosophila
C 31	47	2.9	973	17	CNS008OF	AL052322 Drosophila
C 32	47	2.9	997	17	CNS006GN	AL065132 Drosophila
C 33	46.8	2.8	431	9	AL513947	AL513947
C 34	46.8	2.8	976	17	CNS0068V	AL063041 Drosophila
C 35	46.6	2.8	978	17	AG030617	AG030617 Pan trogl
C 36	46.6	2.8	1101	17	CNS00B3U	AL056389 Drosophila
C 37	46.2	2.8	527	9	AL514325	AL514325
C 38	46.2	2.8	875	17	AG043475	AG043475 Pan trogl
C 39	46.2	2.8	1136	14	BQ943816	BQ943816 AGENCOURT
C 40	46.2	2.8	1417	14	BME10236	BME10236 AGENCOURT
C 41	46	2.8	758	17	AG073529	AG073529 Pan trogl
C 42	45.8	2.8	203	10	AW150511	AW150511 xg32h07.x
C 43	45.8	2.8	471	10	BE503803	BE503803 hv83a05.x
C 44	45.8	2.8	523	9	AL514015	AL514015
C 45	45.8	2.8	766	17	AG041031	AG041031 Pan trogl

ALIGNMENTS

RESULT 1
AW748338/c 289 bp mRNA linear EST 28-APR-2000
LOCUS RC6-BT0252-271099-012-c10 BT0252 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW748338
ACCESSION AW748338
VERSION AW748338.1 GI:7663270
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 289)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zagro, M.A., Bordin, S., Costa, F., Goldmann, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed

Sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC6&t2=RC6-BT0252-271099-012-c10&t3=1998-10-27&t4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 146
 High quality sequence stop: 289.
 Location/Qualifiers
 1. 289
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT0252"
 /dev_stage="Adult"
 /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORES-PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." 74 t

BASE COUNT 94 a 70 c 51 g 74 t
 ORIGIN
 Query Match 14.9%; Score 245.4; DB 10; Length: 289;
 Best Local Similarity 96.9%; Pred. No. 6.3e-39;
 Matches 282; Conservative 0; Mismatches 6; Indels 3; Gaps 3;
 QY 678 GCTGCACCTCACAATGATACAACTAAATACAGTCTCTGTTTATACAGGGAGG 737
 Db 289 GCTGCACCTCACAATGATACAACTAAATACAGTCTCTGTTTATACAGGGAGG 230
 QY 738 CTGATCAATATAATGAATCAAAAGGGGCTGGTCCCATATTGCTGTTTGTGTTG 797
 Db 229 CTGATCAATATAATGAATCAAAAGGGGCTGGTCCCATATTGCTGTTTGTGTTG 171
 QY 798 TTGTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 857
 Db 170 TTGTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 112
 QY 858 GATGTCCTCTCGGGTCTCTGAGGGACCTGGGGAGCTCAGGCTGGGAATCTCCAGGCA 927
 Db 111 GATGTCCTCTCGGGTCTCTGAGGGACCTGGGGAGCTCAGGCTGGG-ATCTCCAGGCA 53
 QY 918 GTAGTCCCTATCAAAATCAAACTCCAGTTCTGGGGGGGAAACAAA 968
 Db 52 GTAGTCCCTATCAAAATCAAACTCCAGTTCTGGGGGGGAAACAAA 2

RESULT 2
 AA527406
 LOCUS n237c02.s1 NCI_CGAP_C03 Homo sapiens cDNA clone IMAGE:936962 3',
 DEFINITION mRNA, sequence.
 AA527406
 VERSION AA527406.1 GI:2269475
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 467)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,
 Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 1057 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 380.
 Location/Qualifiers
 1. 467

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NCI_CGAP_C03"
 /sex="pooled"
 /tissue_type="colon"
 /lab_host="DH10B"
 /note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization." 111 t

BASE COUNT 72 a 113 c 166 g 111 t
 ORIGIN
 Query Match 5.8%; Score 111.4; DB 9; Length 467;
 Best Local Similarity 99.1%; Pred. No. 3e-12;
 Matches 112; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 816 TTGTGGGCTCTCTCTCAATTATGAAGAGGAGCAGTAAAGATGTTCTCTCGGGTCC 875
 Db 1 TTGTGGGCTCTCTCTCAATTATGAAGAGGAGCAGTAAAGATGTTCTCTCGGGTCC 60
 QY 876 TCTCAGGACCTGGGGAGCTCAGGCTGGGATCTCCAAAGGAGTAGTGCCT 928
 Db 61 TCTCAGGACCTGGGGAGCTCAGGCTGGGATCTCCAAAGGAGTAGTGCCT 113

RESULT 3
 AUI35588
 LOCUS PLACE1 Homo sapiens cDNA clone PLACE1002437 5', mRNA
 DEFINITION
 sequence.
 AUI35588
 VERSION AUI35588.1 GI:10956127
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 736)
 Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
 Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
 Isogai,T.
 HRI human cDNA project
 Unpublished (2000)
 Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kiserazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.
 Location/Qualifiers
 1. 736

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="PLACE1002437"

```

/clone_lib="PLACE1"
/tissue_type="placenta"
/notes="Vector: pME18SFL3"
BASE COUNT 163 a 199 c 199 g 170 t 5 others
ORIGIN
Query Match 5.4%; Score 89.4; DB 9; Length 736;
Best Local Similarity 98.9%; Pred. No. 5.3e-08;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1553 GTAAATGCGAGCGAGTGTAGTGGCGGGACCCGAGCGAGCGAGCCACCTTCTCTC 1612
Dbb 1 GTAGTGTGCGAGCGAGTGTAGTGGCGGGACCCGAGCGAGCGAGCCACCTTCTCTC 60

QY 1613 CCGCTCGGCGAGCGAGCGAGCGAGCGAGCTC 1643
Dbb 61 CCGGCTCGGCGAGCGAGCGAGCGAGCGAGCTC 91

RESULT 4
AL698654 547 bp mRNA linear EST 21-MAR-2002
LOCUS DKFZp686N12109_r1 686 (synonym: hicc3) Homo sapiens cDNA clone
DEFINITION DKFZp686N12109 5', mRNA sequence.
ACCESSION AL698654
VERSION AL698654
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 547)
AUTHORS Duesterhoeft,A., Lauber,J., Mewes,H.W., Cassenhuber,J. and Wiemann
,S.
EST (Duesterhoeft, et al.)
JOURNAL Unpublished (1999)
COMMENT Contact: Duesterhoeft A
MIPS
Am Klopferpitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Olagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.
No sl sequence available.
This clone (DKFZp686N12109) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
location/Qualifiers
1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp686N12109"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/notes="Vector: pTriplex2; Site_1: SfiI; Site_2: SfiI;
cDNA collection"
BASE COUNT 112 a 154 c 157 g 124 t
ORIGIN
Query Match 5.4%; Score 89; DB 9; Length 547;
Best Local Similarity 100.0%; Pred. No. 8.5e-08;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1555 AATTGCGAGCGAGTGTAGTGGCGGGACCCGAGCGAGCGAGCCACCTTCTCTCC 1614
Dbb 2 AATTGCGAGCGAGTGTAGTGGCGGGACCCGAGCGAGCGAGCCACCTTCTCTCC 61

QY 1615 GGGCTCGGCGAGCGAGCGAGCGAGCTC 1643
Dbb 62 GGGCTCGGCGAGCGAGCGAGCGAGCTC 90

/clone_lib="MAMMAL"
/tissue_type="mammary gland"
/notes="Vector: pME18SFL3"
BASE COUNT 137 a 205 c 260 g 158 t 3 others
ORIGIN
Query Match 5.4%; Score 89; DB 9; Length 763;
Best Local Similarity 100.0%; Pred. No. 7.5e-08;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1555 AATTGCGAGCGAGTGTAGTGGCGGGACCCGAGCGAGCGAGCCACCTTCTCTCC 1614
Dbb 1 AATTGCGAGCGAGTGTAGTGGCGGGACCCGAGCGAGCGAGCCACCTTCTCTCC 60

QY 1615 GGGCTCGGCGAGCGAGCGAGCGAGCTC 1643
Dbb 61 GGGCTCGGCGAGCGAGCGAGCGAGCTC 89

RESULT 6
AL698654 292 bp mRNA linear EST 14-NOV-1994
LOCUS HSC1ZB081 normalized infant brain cDNA Homo sapiens cDNA clone
DEFINITION c-12b08, mRNA sequence.
ACCESSION 244377
VERSION 244377
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 292)
AUTHORS Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
,M.D., Duprat,S., Houllgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
Sebastiani-Kabaktchis,C. and Tessier,A.
IMAGE: molecular integration of the analysis of the human genome
TITLE
Source
1..763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MAMMAL1000851"
/clone_lib="MAMMAL"
/tissue_type="mammary gland"
/notes="Vector: pME18SFL3"

```

JOURNAL
MEDLINE
COMMENT

and its expression
C. R. Acad. Sci. III, Sci. Vie 378 (2), 263-272 (1995);
95277534
Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_library_id: C; Genexpress_sequence_id: ylc-1zb08
Seq primer: (-21)M13_universal.

FEATURES

source

```
1..292
Location/Qualifiers
  1..292
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="C-1zb08"
    /clone_lib="normalized infant brain cDNA"
    /sex="Female"
    /tissue_type="total brain"
    /dev_stage="3 months old"
    /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
    Site_2: NotI; sex:female; dev_stage=3 months old;
    isolate-muscular atrophy patient; tissue_type=total brain
    ; total mRNA was oligo-(dt) primed and directionally
    cloned 5' -> 3' into the HindIII -> NotI sites of the
    lafmid BA vector. Clone library from B Soares, Psychiatry
    Dept. Columbia University, USA. Normalization method:
    Bento Soares, P.N.A.S in press"
```

BASE COUNT

ORIGIN

```
50 a 87 c 96 g 56 t 3 others
```

```
Query Match 4.6%; Score 75; DB 14; Length 292;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1569 GTGAGTGGGGCGGGACCCGAGAGCCGAGCCGACCCCTCTCTCCGGGCTCGGCAGGG 1628
```

```
|||||
Db 1 GTGAGTGGGGCGGGACCCGAGAGCCGAGCCGACCCCTCTCTCCGGGCTCGGCAGGG 60
```

```
Qy 1829 CAGGCGGGGAGCTC 1643
```

```
|||||
Db 61 CAGGCGGGGAGCTC 75
```

```
RESULT 7
BC034824
LOCUS BC034824 Homo sapiens, clone IMAGE:4749735, mRNA.
DEFINITION BC034824 Homo sapiens, clone IMAGE:4749735, mRNA.
ACCESSION BC034824.1 GI:21961558
VERSION BC034824.1 GI:21961558
KEYWORDS HTC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1004)
Direct Submission
Strausberg, R.
Submitted (24-JUL-2002) National Institutes of Health, Mamalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
```

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps-femail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (JLNLI)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Günaratne, P.H., Garcia, A.M., Lu, X., Kulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/JLNLI at: <http://image.llnl.gov>
Series: iRAK Plate: 52 Row: d Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5915657
This clone has the following problem: incomplete processing.

FEATURES

source

```
1..1004
Location/Qualifiers
```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4749735"
/tissue_type="Skin, squamous cell carcinoma"
/clone_lib="NCI_CGAP_Skn4"
/lib_host="DH10B"
/note="Vector: pCMV-SPORT6.ccdB"
```

```
306 a 234 c 230 g 234 t
```

BASE COUNT

ORIGIN

```
Query Match 4.6%; Score 75; DB 11; Length 1004;
Best Local Similarity 100.0%; Pred. No. 4.3e-05;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1569 GTGAGTGGGGCGGGACCCGAGAGCCGAGCCGACCCCTCTCTCCGGGCTCGGCAGGG 1628
```

```
|||||
Db 1 GTGAGTGGGGCGGGACCCGAGAGCCGAGCCGACCCCTCTCTCCGGGCTCGGCAGGG 60
```

```
Qy 1629 CAGGCGGGGAGCTC 1643
```

```
|||||
Db 61 CAGGCGGGGAGCTC 75
```

RESULT 8

CNS015W3

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CNS015W3 1201 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN15123 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL105981
AL105981.1 GI:5619397
GSS.
Drosophila melanogaster.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1201)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk> . This Drosophila melanogaster BAC
library (DrosBAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
peloBAC11.
Location/Qualifiers
1..1201
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN15123"
/clone_lib="DrosBAC"
/p_asmid="pBelobAC11"
/note="end : SP6"


```

Qy 1360 TCGTGGCGGCTGACGTCG 1379
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 262 GGGGGGGGGGGGGGGGGCG 281

RESULT 11
CNS0052P      844 bp      DNA      linear      GSS 03-JUN-1999
LOCUS
DEFINITION   Drosophila melanogaster genome survey sequence TET3 end of BAC #
              BACR11p16 of RPI-98 library from Drosophila melanogaster (fruit
              fly); genomic survey sequence.
ACCESSION    AL056652.1   GI:4932342
VERSION      AL056652.1   GI:4932342
KEYWORDS     GSS.
SOURCE      Drosophila melanogaster.
ORGANISM    Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 844)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
            p1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             Location/Qualifiers
     source           1..844
                     /organism="Drosophila me-anogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR11p16"
                     /clone_lib="RPI-98"
                     /note="end : TET3"

BASE COUNT          261 a 112 c 92 g 35 t 344 others

Query Match
Best Local Similarity 3.3%; Score 53.8; DB 17; Length 844;
Matches 86; Conservative 131; Mismatches 140; Indels 5; Gaps 1;

Qy 1117 AAACTAACAAGGAAAAAATTCGGGAAGCAGCATTTACAGGAACAATCCACTG 1176
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 301 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAA 360

Qy 1177 GTGCCCTTGGTCGGGGAAGCTGGACTAGAGTCTCGGGCGAGCCCGGACCGCG 1236
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 361 SASVSSACSSVAGSMAAGASCBGCAANMCAACCCSSSSASRSRMSSSSSSSSS 420

Qy 1237 CTTCCGGCGCTCTTAGCGCGCGCGCGCGCGCGGGAAGGGAGCGAGCCGGGACC 1296
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 421 VGSSSCSAMSSCCCVSSCGMASSCGSGMSASSSSSSSSSSSSSSVSGVAGRGCMCC 480

Qy 1297 CTAAGACACTCTCTACCTCCACCCACCCACCCACCTCCCCCACTCCCTAGAT 1356
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 481 S-----MCCCMCMSCMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCMCM 535

Qy 1357 GTGTCGGCGGCGCTGACGTCGCTTTAAGCGGGGGCGCGCGCTCCACCTGCTTTC 1416
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 536 SGRGGSSSVSSGVSSSVGVGACACSSASSVSSSSSVSSSVSSSVSSSVSSSVSS 595

```

```

Qy 1417 TCGTGAAGTGAACGACTACATAACAGACGCGCGGGAACGCGGGGAGGAGAGCA 1476
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 596 GGGSRVSSCGSSSSSSSSGCVSRSRSGSANGVRVYGGSGRGGGRRGGANARAD 655

Qy 1477 CA 1478
|
Db 656 DA 657

RESULT 12
AU222489      237 bp      mRNA      linear      EST 03-APR-2002
LOCUS
DEFINITION   AU222489 Rice callus (2001) Oryza sativa (japonica cultivar-group)
              cDNA clone C10354, mRNA sequence.
ACCESSION    AU222489
VERSION      AU222489.1   GI:15008101
KEYWORDS     EST.
SOURCE      Oryza sativa (japonica cultivar-group).
ORGANISM    Oryza sativa (japonica cultivar-group).
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE    1 (bases 1 to 237)
AUTHORS     Sasaki, T. and Yamamoto, K.
TITLE       Rice cDNA from callus (2001)
JOURNAL     Unpublished (2001)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
            305-8602, Japan
            Tel: 81-298-38-7441
            Fax: 81-298-38-7468
            Email: tsasaki@nri.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
            PROJECT = "RGP".

FEATURES             Location/Qualifiers
     source           1..237
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /clone="C10354"
                     /clone_lib="Rice callus (2001)"
                     /tissue_type="callus"

BASE COUNT          69 a 24 c 38 g 102 t 4 others

Query Match
Best Local Similarity 3.2%; Score 53; DB 9; Length 237;
Matches 86; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 716 CTGTGTTTTCACAGGGAGGCTGATCAATATATGAATATAAGGGGCTGTCCCA 775
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 17 CTTAATTTAATTCGAAGNGTGTNTTNGCAAAAAAAGGAGGCGCGCCCTT 76

Qy 776 TATTCCTTCGTGTTTGTTCGTTTGTTCCTTTTGTTCCTTCCTTCCTTCCTC 835
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 77 TTTTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCT 136

Qy 836 AATTATCAAGAGACAGACIAGA 859
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 137 CACTGGGAAGGGGAGAGATTGA 160

RESULT 13
CNS0060N/c
LOCUS
DEFINITION   CNS0060N Drosophila melanogaster genome survey sequence 17 end of BAC #
              BACR14J21 of RPI-98 library from Drosophila melanogaster (fruit
              fly); genomic survey sequence.
ACCESSION    AL065629
VERSION      AL065629.1   GI:4944698
KEYWORDS     GSS.
SOURCE      Drosophila melanogaster.

```



```

ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 910)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr)
          Determination of this BAC-end sequence was carried out as part of a
          collaboration with the Berkeley Drosophila Genome Project (BDGP).
          The BDGP is constructing a physical map of the Drosophila
          melanogaster genome using these BACs. For further information
          please see http://www.fruitfly.org The BDGP Drosophila
          melanogaster BAC library was prepared by Kazutoyo Osoegawa and
          Aaron Mamoser in Peter de Jong's laboratory in the Department of
          Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
          NY. The library is named RPCI-98 and was constructed by partial
          EcoRI digestion of Drosophila DNA provided by the BDGP from the
          P1 and EST libraries. A more detailed description of the library
          and how to order individual BAC clones, the entire library, or
          filters for hybridization from the BACPAC Resource Center can be
          found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES             Location/Qualifiers
     source          1..910
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACR14721"
                     /clone_lib="RPCI-98"
                     /note="end : 17"
BASE COUNT 202 a 63 c 112 g 138 t 335 others
ORIGIN
Query Match          3.2%; Score 52.4; DB 17; Length 910;
Best Local Similarity 20.5%; Pred. No. 1.4;
Matches 48; Conservative 104; Mismatches 82; Indels 0; Gaps 0;
QY 1118 AACTAACAAAGAAATAATTCGGGAAGCAGGATTTAGAGGAAGCAATTCCTACTCG 1177
      ||:||||: ||:||||: ||:||||: ||:||||: ||:||||: ||:||||: ||:||||:
Db 752 AASMARVAAVARVASAAVAVASAAASASASASASASASASASASASASASASAS 693
QY 1178 TCCCTTGCTCGCGGACGTTGACTAGAGAGTCTGGCGGCGACGCCGAGCCAGCGC 1237
      ||:||||: ||:||||: ||:||||: ||:||||: ||:||||: ||:||||: ||:||||:
Db 692 SASSSASASASASASASASASASASASASASASASASASASASASASASASAS 533
QY 1238 TTCCTCGCGGCTTTAGCGCGCGGCGCGCGGCGGAGAGCGACGACCGCGGACCC 1297
      ||:||||: ||:||||: ||:||||: ||:||||: ||:||||: ||:||||: ||:||||:
Db 632 SCSSRSGCSASSSASASSSSSSSSSSSSSSSCGCGSCASCSSSSRSGCGGSSVGS 573
QY 1298 TAAGACACTGCTGACCTCCACCCACCCACCCACCTCCCTCCCTCCCTCCCTCC 1351
      ||:||||: ||:||||: ||:||||: ||:||||: ||:||||: ||:||||: ||:||||:
Db 572 SSCGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 519

RESULT 14
CNS015XC          832 bp      DNA      linear      GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC
DEFINITION BACN15808 of DrosBAC library from Drosophila melanogaster (fruit fly); genomic survey sequence.
ACCESSION AL106026
VERSION AL106026.1 GI:5613635
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
          Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
          Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
          Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 832)
AUTHORS Genoscope.
TITLE Direct Submission

JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
COMMENT - Web : www.genoscope.cns.fr)
          Determination of this BAC-end sequence was carried out as part of a
          collaboration with the European Drosophila Genome Project (EDGP) -
          http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
          library (Dros BAC) was made by Alain Billaut at CEPH (Centre
          d'Etude de Polymorphisme Humain) with funding provided by a MRC
          project grant. The DNA was prepared from embryos by Alain Bucheton
          and Genevieve Payan. It has been constructed in the vector
          pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1..832
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone="BACN15B08"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : T7"
BASE COUNT 177 a 118 c 105 g 184 t 248 others
ORIGIN
Query Match          3.1%; Score 51.2; DB 17; Length 832;
Best Local Similarity 38.6%; Pred. No. 2.6;
Matches 93; Conservative 42; Mismatches 106; Indels 0; Gaps 0;
QY 1125 AAGGAAAAAATAITGCGGAAGCAGGATTTAGAGGAAGCAATTCCTACTGGTCCCTT 1184
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 142 AAAAAAANAANTANNANATGATGAGGAGGATVATRAAAVAHAHVMMVVVSSSSSG 201
QY 1185 GGCTCGCGGGAACGTGACTAGAGAGTCTGGCGGCGAGCCCGAGCCCGAGCGCTTCCGCG 1244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 202 GGGGCGGGGCGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 261
QY 1245 GCGTCTTAGCGCGCGCGCGCGCGCGGGGAGGGAGGAGCCAGCGCGACCTAAGACA 1304
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 262 CTCCTATTNATGGTGGGGTGTGTTTGTGAGTGGTGGTGGTGGTGGTGGTGGTGG 321
QY 1305 CCTGTGTATCTCCACCCCGCCACCCCGCCACCCCGCCACCCCGCCACCCCGCC 1364
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 322 CCCCMMYCCCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 381
QY 1365 G 1365
Db 382 G 382

RESULT 15
CNS02FE9          581 bp      DNA      linear      GSS 13-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
DEFINITION 132116 of library G from Tetraodon nigroviridis, genomic survey
          sequence.
ACCESSION AL194922
VERSION AL194922.1 G:7833028
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
          Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
          Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 581)
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,I., Fisher,C.,
          Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
          Saurin,W. and Weissenbach,J.
          Human gene number estimate provided by genome wide analysis using
          Tetraodon nigroviridis DNA sequence
          Unpublished
          2 (bases 1 to 681)
          Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
          Bouneau,I., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
          Weissenbach,J.

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CorpuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 12:46:48 ; Search time 448.725 Seconds
(without alignments)
8245.650 Million cell updates/sec

Title: US-09-596-141C-3

Perfect score: 1643

Sequence: 1 gaattcctgctggtgctc.....caggcgaggcgaggagctc 1643

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapex 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1643	100.0	1643	22 AAF24681	Nucleotide sequenc
2	1643	100.0	1643	22 AAF24703	Nucleotide sequenc
3	1498.6	91.2	183999	22 AAF28231	Human ABC1 genomic
4	1485.4	90.4	3231	24 AAD37265	Human ABC1 transcr
5	1411.4	85.9	2910	24 AAD37267	Human ABC1 transcr
6	1146.4	69.8	1197	24 ABL38400	Human large ATP-bi
7	119.4	7.3	227	21 AAC09615	Human secreted pro
8	102.2	6.2	2910	24 AAD37267	Human ABC1 transcr
9	102.2	6.2	3231	24 AAD37265	Human ABC1 transcr

10	98	6.0	7260	22 AAD21326	Human ATP binding
11	98	6.0	7260	22 AAI70315	Human ATP binding
12	97	5.9	1643	22 AAF24681	Nucleotide sequenc
13	97	5.9	1643	22 AAF24703	Nucleotide sequenc
14	92.6	5.6	183999	22 AAF28231	Human ABC1 genomic
15	91	5.5	221	24 AAD37267	Human ABC1 gene ex
16	89.4	5.4	736	22 AAD37267	Human ABC1 gene ex
17	89.4	5.4	1556	22 AAI8606	Human cDNA clone (
18	89	5.4	763	22 AAI8606	Human cDNA clone (
19	89	5.4	1750	22 AAI8606	Human cDNA clone (
20	77.2	4.7	7281	22 AAI8606	Human cDNA clone (
21	75.4	4.6	7086	22 AAI8606	Human cDNA clone (
22	75.4	4.6	7086	22 AAI8606	Human cDNA clone (
23	75	4.6	9854	22 AAI8606	Human cDNA clone (
24	67	4.1	10442	22 AAI8606	Human cDNA clone (
25	67	4.1	10442	22 AAI8606	Human cDNA clone (
26	60	3.7	10474	22 AAI8606	Human cDNA clone (
27	60	3.7	10474	22 AAI8606	Human cDNA clone (
28	60	3.7	10474	22 AAI8606	Human cDNA clone (
29	60	3.7	10474	22 AAI8606	Human cDNA clone (
30	51.4	3.1	474	24 AAI8606	Human cDNA clone (
31	46.6	2.8	5116	24 AAI8606	Human cDNA clone (
32	45	2.7	114955	20 AAI8606	Human cDNA clone (
33	44.2	2.7	456	22 AAI8606	Human cDNA clone (
34	43.8	2.7	385	22 AAI8606	Human cDNA clone (
35	43.8	2.7	400	22 AAI8606	Human cDNA clone (
36	43.4	2.6	401	22 AAI8606	Human cDNA clone (
37	43.4	2.6	772	22 AAI8606	Human cDNA clone (
38	43.4	2.6	2339	22 AAI8606	Human cDNA clone (
39	43.4	2.6	10119	22 AAI8606	Human cDNA clone (
40	43.4	2.6	14983	22 AAI8606	Human cDNA clone (
41	43.4	2.6	14983	22 AAI8606	Human cDNA clone (
42	43.4	2.6	32190	22 AAI8606	Human cDNA clone (
43	43.4	2.6	32249	22 AAI8606	Human cDNA clone (
44	43.4	2.6	39068	22 AAI8606	Human cDNA clone (
45	43.4	2.6	39068	22 AAI8606	Human cDNA clone (

ALIGNMENTS

RESULT 1

AAF24681
ID AAF24681 standard; DNA; 1643 BP.

XX AC AAF24681:

XX 20-APR-2001 (first entry)

XX Nucleotide sequence of the 5' flanking region of the human ABC1 gene.

XX Human; adenosine triphosphate binding cassette protein 1; ABC1;

XX apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;

XX chromosome 9q22-9q31; heart disease; hypercholesterolemia;

XX atherosclerosis; cholesterol transport; ss.

XX Homo sapiens.

XX WO2000078972-A2.

XX PD 28-DEC-2000.

XX PF 16-JUN-2000; 2000WO-US16765.

XX PR 18-JUN-1999; 99US-0140264.

XX PR 14-SEP-1999; 99US-0153872.

XX PR 19-NOV-1999; 99US-0166573.

XX PA (CVTH-) CV THERAPEUTICS INC.

XX PI Lawn RM, Wade D, Garvin M;

XX WP; 2001-137812/14.

01/11/00
01/11/00

XX
AC AAF24703;
XX
XX 20-A2R-2001 (first entry)
XX
XX Nucleotide sequence of the 5' flanking region of the human ABC1 gene.
XX
XX Human; adenosine triphosphate binding cassette protein 1; ABC1;
KW apolipoprotein-mediated mobilisation; cholesterol; tangier disease;
KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
KW atherosclerosis; cholesterol transport; ss.
XX
XX Homo sapiens.
XX
XX W0200078971-A2.
XX
XX 28-DEC-2000.
XX
XX 16-JUN-2000; 2000WO-US16591.
XX
XX 18-JUN-1999; 99US-0140264.
PR 14-SEP-1999; 99US-0153872.
PR 19-NOV-1999; 99US-0166573.
XX
XX (CVTH-) CV THERAPEUTICS INC.
XX (UNIW) UNIV WASHINGTON.
PI Lawn RM, Wade D, Oram JF, Garvin M;
XX WPI; 2001-137811/14.
XX
XX Adenosine triphosphate (ATP) binding cassette protein (ABC) 1
PT polynucleotides and polypeptides, useful for treatment of heart disease
PT and other disorders associated with hypercholesterolemia and
PT atherosclerosis -
XX
XX Disclosure; Page 138-139; 21pp; English.
XX
XX The present sequence represents the 5' flanking region of the human:
CC adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABC1
CC resides in cell membranes and utilizes ATP hydrolysis to transport a wide
CC variety of substrates across the plasma membrane. ABC1 is a pivotal
CC protein in the apolipoprotein-mediated mobilisation of intracellular
CC cholesterol stores. ABC1 is defective in Tangier disease, a genetic
CC disorder characterised by abnormal HDL-cholesterol metabolism. The ABC1
CC gene is localised to chromosome 9q22-9q31. The ABC1 genes and proteins
CC are useful for developing pharmaceutical agents for the treatment of
CC heart disease and other disorders associated with hypercholesterolemia
CC and atherosclerosis. The genes are useful for developing screening assays
CC to screen for compounds that regulate the expression of genes associated
CC with cholesterol transport. The genes and proteins are also useful for
CC are also useful as diagnostic indicators of cardiovascular disease and
CC other disorders associated with hypercholesterolemia.
XX
SQ Sequence 1643 BP; 370 A; 413 C; 457 G; 403 T; 0 other;
Query Match 100.0%; Score 1643; DB 22; Length 1643;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1643; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCCTTGTGCTGGCTCCACATGCACTTCCAGGGCCCTGCTTGTGCTCTTCTATGGGTC 60
DB 1 GAATTCCTTGTGCTGGCTCCACATGCACTTCCAGGGCCCTGCTTGTGCTCTTCTATGGGTC 60
QY 61 TGTCTTGTGCTTGTATGAACCACTGATGTGAGTACCTGGGCTTGAGGCTGGCTGGAGA 120
DB 61 TGTCTTGTGCTTGTATGAACCACTGATGTGAGTACCTGGGCTTGAGGCTGGCTGGAGA 120
QY 121 TCTCTTGTGCTTGTATGAACCACTGATGTGAGTACCTGGGCTTGAGGCTGGCTGGAGA 180
DB 121 TCTCTTGTGCTTGTATGAACCACTGATGTGAGTACCTGGGCTTGAGGCTGGCTGGAGA 180
QY 181 AGTCTGGAATATGATGGAGCTGGAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240

DB 181 AGTCTGGAATATGATGGAGCTGGAGGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 241 ATGCCACCTCATTTCTGGCCAAAACCTCAGGTCAAACCTGTGAAGAGCTTAAATGTGAATCTG 300
DB 241 ATGCCACCTCATTTCTGGCCAAAACCTCAGGTCAAACCTGTGAAGAGCTTAAATGTGAATCTG 300
QY 301 CCCTTCAGGTGGCTACAAAGGATCTCTTGTCAAGGTAGGAGACCTTGTGGCCTCCACGT 360
DB 301 CCCTTCAGGTGGCTACAAAGGATCTCTTGTCAAGGTAGGAGACCTTGTGGCCTCCACGT 360
QY 361 GCACCTCCAGGGCCTGCTTGGCCCTCTTCTACGGGTCTGTCTGAGTCTTCTATGAATCTC 420
DB 361 GCACCTCCAGGGCCTGCTTGGCCCTCTTCTACGGGTCTGTCTGAGTCTTCTATGAATCTC 420
QY 421 CTTTCAGGGCAGATCATATTTAGACTCTTACAGTGTGACCTGAGTTTGGCCAGAGATA 480
DB 421 CTTTCAGGGCAGATCATATTTAGACTCTTACAGTGTGACCTGAGTTTGGCCAGAGATA 480
QY 481 AGGTGACATTTAGTTTGTGGCTTCATCAATGACTTAAATTTAGACATATGCTGTGTA 540
DB 481 AGGTGACATTTAGTTTGTGGCTTCATCAATGACTTAAATTTAGACATATGCTGTGTA 540
QY 541 GGCCTGCATTCCTACTCTTGGCTTTTGTGGCCCTCCAGTGTGTTGGGTAGTTTGTCT 600
DB 541 GGCCTGCATTCCTACTCTTGGCTTTTGTGGCCCTCCAGTGTGTTGGGTAGTTTGTCT 600
QY 601 CCCCTACAGCCAAAGGCAACAGATAAGTTGGAGGCTTGGAGTGGCTGACATATTTTAC 660
DB 601 CCCCTACAGCCAAAGGCAACAGATAAGTTGGAGGCTTGGAGTGGCTGACATATTTTAC 660
QY 661 ACAGCTCAATTTCTGCTGCTCACTTCACAAATGATATCAAACTAAATACAGTCCGTGTG 720
DB 661 ACAGCTCAATTTCTGCTGCTCACTTCACAAATGATATCAAACTAAATACAGTCCGTGTG 720
QY 721 TTTTATACAGGGAGGCTGATCAATATATGAATTAATGAATTAATGAATTAATGAATTAATGA 780
DB 721 TTTTATACAGGGAGGCTGATCAATATATGAATTAATGAATTAATGAATTAATGAATTAATGA 780
QY 781 TTTCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 840
DB 781 TTTCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 840
QY 841 ATGAAGAGAGACGCTAGATCTTCTCTCGGGTCTCTCTGAGGGACCTGGGGAGCTCAGGC 900
DB 841 ATGAAGAGAGACGCTAGATCTTCTCTCGGGTCTCTCTGAGGGACCTGGGGAGCTCAGGC 900
QY 901 TGGGAATCTCCAAAGGCACTAGTTCGCTTATCAAAATCAAAAGTCCAGTTTGTGGGGGA 960
DB 901 TGGGAATCTCCAAAGGCACTAGTTCGCTTATCAAAATCAAAAGTCCAGTTTGTGGGGGA 960
QY 961 AAACAAG 1020
DB 961 AAACAAG 1020
QY 1021 TTTGAAG 1080
DB 1021 TTTGAAG 1080
QY 1081 AGCTCTCTCTCCCAATCCCTCCCTCCGCTGAGGAACTAAACAAGAGAGAGAGAGAGAGAGAG 1140
DB 1081 AGCTCTCTCTCCCAATCCCTCCCTCCGCTGAGGAACTAAACAAGAGAGAGAGAGAGAGAGAG 1140
QY 1141 CGGAAAG 1200
DB 1141 CGGAAAG 1200
QY 1201 GACTAG 1260
DB 1201 GACTAG 1260
QY 1261 GCGCGGGGGAG 1320


```

1999 TCTCTGGCTGCATCTCACAAGATATACAACTAAATACAAAGTCCCTGTTCTTATCAC 2058
731 AGGGAGGTGATCAATATAATCAAAATTAAGAGGGGCTGGTCCCATATGTTCTGTGTT 790
2059 AGGGAGGTGATCAATATAATCAAAATTAAGAGGGGCTGGTCCCATATGTTCTGTGTT 2117
791 TTG-----TTGTTGTTGTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 841
2118 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2177
842 TGAAGAGAGAGCAGTAAATGATGTCCTCTCGGTCCTCTGAGGAGCCTGGGAGCTCAGGCT 901
2178 TGAAGAGAGAGCAGTAAATGATGTCCTCTCGGTCCTCTGAGGAGCCTGGGAGCTCAGGCT 2237
902 GGAATCTCCAGGAGCAGTAAATGATGTCCTCTCGGTCCTCTGAGGAGCCTGGGAGCTCAGGCT 961
2238 GGAATCTCCAGGAGCAGTAAATGATGTCCTCTCGGTCCTCTGAGGAGCCTGGGAGCTCAGGCT 2297
962 AACAAAGCAGCCCATACCCAGAGAGCTGTCGGCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1021
2298 AACAAAGCAGCCCATACCCAGAGAGCTGTCGGCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 2357
1022 TTGAAGAGAGAGCAGTAAATGATGTCCTCTCGGTCCTCTGAGGAGCCTGGGAGCTCAGGCT 1081
2358 TTGAAGAGAGAGCAGTAAATGATGTCCTCTCGGTCCTCTGAGGAGCCTGGGAGCTCAGGCT 2417
1082 GCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1141
2418 GCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2477
1142 CGGAAGAGAGAGTAAAGAGAGAGAGTAAAGAGAGAGTAAAGAGAGAGTAAAGAGAGAGTAAAG 1201
2478 CGGAAGAGAGAGTAAAGAGAGAGAGTAAAGAGAGAGTAAAGAGAGAGTAAAGAGAGAGTAAAG 2537
1202 ACTAGAGAGTCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1261
2538 ACTAGAGAGTCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2597
1262 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1321
2598 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2657
1322 -----CCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1376
2658 CCGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2717
1377 TCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1436
2718 TCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2777
1437 TAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1496
2778 TAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2837
1497 CCGTCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1556
2838 CCGTCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2897
1557 TCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1616
2898 TCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2957
1617 CCGTCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1643
2958 GCCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2984

RESULT 5
AAD37267
ID AAD37267 standard; DNA; 2910 BP.
XX
AC AAD37267;
```

```

XX 21-AUG-2002 (first entry)
DT Human ABC1 transcription regulatory DNA #3.
XX
DE Human; ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
XX cholesterol metabolism; hypercholesterolaemia; antisense therapy; ds.
KW
XX Homo sapiens.
OS
XX
FH Key
FH protein_bind
FT 12..23
FT /tag= a
FT /bound_moiety= "LMO2C04/WYCD"
FT
FT protein_bind
FT 97..107
FT /tag= b
FT /bound_moiety= "DeltaEF1"
FT
FT protein_bind
FT 110..125
FT /tag= c
FT /bound_moiety= "SB/NKX2.5"
FT
FT protein_bind
FT 196..211
FT /tag= d
FT /bound_moiety= "S8"
FT
FT protein_bind
FT 328..237
FT /tag= e
FT /bound_moiety= "GATA"
FT
FT protein_bind
FT 399..420
FT /tag= f
FT /bound_moiety= "IK2"
FT
FT protein_bind
FT 412..420
FT /tag= g
FT /bound_moiety= "LYF1"
FT
FT protein_bind
FT 528..539
FT /tag= h
FT /bound_moiety= "LMO2C04/WYCD/DeltaEF1"
FT
FT protein_bind
FT 549..556
FT /tag= i
FT /bound_moiety= "LYF1"
FT
FT protein_bind
FT 558..568
FT /tag= j
FT /bound_moiety= "DeltaEF1"
FT
FT protein_bind
FT 590..596
FT /tag= k
FT /bound_moiety= "LNX2.5"
FT
FT protein_bind
FT 608..620
FT /tag= l
FT /bound_moiety= "NFY/CAA1"
FT
FT protein_bind
FT 708..715
FT /tag= m
FT /bound_moiety= "MZF1"
FT
FT protein_bind
FT 723..730
FT /tag= n
FT /bound_moiety= "MZF1"
FT
FT protein_bind
FT 771..785
FT /tag= o
FT /bound_moiety= "HPH2/SRY/EV11"
FT
FT protein_bind
FT 803..812
FT /tag= p
FT /bound_moiety= "CREBP1/VBP"
FT
FT protein_bind
FT 831..837
FT /tag= r
FT /bound_moiety= "NKX2.5"
FT
FT protein_bind
FT 1076..1089
FT /tag= s
FT /bound_moiety= "GATA"
FT
FT protein_bind
FT 1173..1188
FT /tag= t
FT /bound_moiety= "LXRalpha/deltaEF1"
FT
FT protein_bind
FT 1189..1209
FT /tag= u
FT /bound_moiety= "DeltaEF1/LYF1/IK2"
FT
FT protein_bind
FT 1483..1491
FT /tag= v
```



```

FT protein_bind /bound_moiety= "AP4"
FT 1498..1514 /tag= w
FT /bound_moiety= "LMO2-COM/MYOD/deltaEFL/E47"
FT 1524..1545 /tag= x
FT /bound_moiety= "ZID/deltaEFL"
FT 1597..1607 /tag= y
FT /bound_moiety= "DeltaEFL"
FT 1622..1627 /tag= z
FT /bound_moiety= "PPAR"
FT 1632..1637 /tag= aa
FT /bound_moiety= "PPAR"
FT 1685..1698 /tag= ab
FT /bound_moiety= "USE/NMYC/MYCKAX"
FT 1787..1797 /tag= ac
FT /bound_moiety= "DeltaEFL"
FT 1809..1819 /tag= ad
FT /bound_moiety= "DeltaEFL"
FT 1822..1833 /tag= ae
FT /bound_moiety= "SRY"
FT 1840..1850 /tag= af
FT /bound_moiety= "AP1"
FT 1942..1956 /tag= ag
FT /bound_moiety= "HNF3beta"
FT 1978..1985 /tag= ah
FT /bound_moiety= "NKX2.5"
FT 2008..2016 /tag= ai
FT /bound_moiety= "PPAR/NKX2.5/PPAR"
FT 2019..2024 /tag= aj
FT /bound_moiety= "PPAR/NKX2.5/PPAR"
FT 2051..2059 /tag= ak
FT /bound_moiety= "GATA"
FT 2104..2111 /tag= al
FT /bound_moiety= "SOX5"
FT 2114..2152 /tag= am
FT /bound_moiety= "SRY/HNF3beta"
FT 2221..2228 /tag= an
FT /bound_moiety= "MZFL"
FT 2234..2249 /tag= ao
FT /bound_moiety= "IK2/NFkappaB/CREL"
FT 2259..2272 /tag= ap
FT /bound_moiety= "LMO2COM/GATA"
FT 2289..2306 /tag= aq
FT /bound_moiety= "MZFL/SRY"
FT 2313..2318 /tag= ar
FT /bound_moiety= "PPAR"
FT 2321..2326 /tag= as
FT /bound_moiety= "PPAR"
FT 2335..2342 /tag= at
FT /bound_moiety= "MZFL"

```

```

FT protein_bind 2361..2384 /tag= au
FT /bound_moiety= "HNF3beta/SRY/EVL"
FT 2426..2433 /tag= av
FT /bound_moiety= "MZFL"
FT 2442..2451 /tag= aw
FT /bound_moiety= "AP4"
FT 2455..2466 /tag= ax
FT /bound_moiety= "SRY"
FT 2491..2498 /tag= ay
FT /bound_moiety= "STAT"
FT 2524..2534 /tag= az
FT /bound_moiety= "STAT/PPAR"
FT 2536..254 /tag= ba
FT /bound_moiety= "PPAR"
FT 2589..2600 /tag= bb
FT /bound_moiety= "AP2"
FT 2610..2617 /tag= bc
FT /bound_moiety= "MZFL"
FT 2634..2648 /tag= bd
FT /bound_moiety= "LMO2COM/MYOD/E47"
FT 2657..2672 /tag= be
FT /bound_moiety= "RRE31"
FT 2680..2698 /tag= bf
FT /bound_moiety= "MZFL/CMYB"
FT 2728..2740 /tag= bg
FT /bound_moiety= "SP1/GC"
FT 2743..2757 /tag= bh
FT /bound_moiety= "USE/NMYC/ARNT"
FT 2758..2773 /tag= bi
FT /bound_moiety= "NFE2A1"
FT 2774..2787 /tag= bj
FT /bound_moiety= "XFDI/HFH"
FT 2794..2806 /tag= bk
FT /bound_moiety= "GC/SP1/MZF1"

```

Query Match 85.9%; Score 1411.4; DB 24; Length 2910;

Best Local Similarity 98.0%; Pred. NO. 0;

Matches 1541; Conservative 0; Mismatches 6; Indels 26; Gaps 10;

QY 16 GCCTCCAGATGCACCTCCAGGGCCCTGCTGG--CCTCTCATGGCTCTCTCTCTGAGTGT 73

DB 1345 GCCTCCAGATGCACCTCCAGGGCCCTGCTGGCCCTCTCTATGCTCTCTCTGAGTGT 1404

QY 74 GATAGAACACCTGATGTGAGTACCTGGGCTTGAG--CGTGGCCTGGAGATCCCTGTGACTG 132

DB 1405 GATAGAACACCTGATGTGAGTACCTGGGCTTGAGCCCTGGAGATCCCTGTGACTG 1464

QY 133 TAGCATGGAGGGGGTGTG--CAGCTCAATGTCTGATGAGGTGGTGGAGTCTGGGAT 191

DB 1465 TAGCATGGAGGGGGTGTG--CAGCTCAATGTCTGATGAGGTGGTGGAGTCTGGGAT 1524

QY 192 ATGATGGAGCTGGAGGTGGAGAGAGTAGGCTGGGGCAGCTCTCTCATGCCACCTCA 251

DB 1525 ATGATGGAGCTGGAGGTGGAGAGAGTAGGCTGGGGCAGCTCTCTCATGCCACCTCA 1584

QY 252 TTCTGGCCCAACACAGGTCAAACCTGTGAAGAGTCTAAATGTGAATCTGCCCTCAAGT 311

```

1595 TTCGGCCAAACTCAGGTCAAACTGTGAAGAGTCTAAAGTGAATCTGCCCTTCAAGGT 1644
QY 312 GGCTCAAAAGTATCTTTGTCAGAGTAGAGACCTTGTGGGCTCCACGTGACTTCCAGG 371
Db 1645 GSCTCAAAAGGTATCTTTGTCAGAGTAGAGACCTTGTGGGCTCCACGTGACTTCCAGG 1704
QY 372 GCCTGCTT-GGCTCTTCTACGGGTCTGTCTGAGTCTTCTATGAATCTCCCTCAGGGC 430
Db 1705 GGCTGCTTGGGCTCTTCTACGGGTCTGTCTGAGTCTTCTATGAATCTCCCTCAGGGC 1761
QY 431 AGATTCAATTTAGACTCTTACAGCTTTCAGCTGAGTCTTGGCCAGAAATAGGTGACATT 490
Db 1762 AGATTCAATTTAGACTCTTACAGCTTTCAGCTGAGTCTTGGCCAGAAATAGGTGACATT 1821
QY 491 TAGTTGTGTGGCTTGATGAATGACTTAATATTTAGACATATGCTGTAGGCTGCATT 550
Db 1822 TAGTTGTGTGGCTTGATGATGACTTAATATTTAGAC-ATGGTGTGTAGSCCTGCATT 1879
QY 551 CTACTCTTGGCTTTTTCCTCCCTCCAGTGTCTTGGGTAGTTCTGCTCCCTACAG 610
Db 1880 CTACTCTTGGCTTTTTCCTCCCTCCAGTGTCTTGGGTAGTTCTGCTCCCTACAG 1938
QY 611 CCAAGGCAAAACAGATAGTTGAGGTCTGGAGTGGCTACATAATTTACAGCACTGCAA 670
Db 1939 CCAAGGCAAAACAGAGAGTTGGAGTCTGGAGTGGCTACATAATTTACAGCACTGCAA 1998
QY 671 TTCTGTGGCTGCATCTCAGAAATGATATACAACTAAATACAAAGTCTGTGTTTATCAC 730
Db 1999 TTCTGTGGCTGCATCTCAGAAATGATATACAACTAAATACAAAGTCTGTGTTTATCAC 2058
QY 731 AGGGAGGCTGATCAATATATCAATATTAAGAGGGGCTGTCCCATATGTCTGTGTT 790
Db 2059 AGGGAGGCTGATCAATATATCAATATTAAGAGGGGCTGTCCCATATGTCTGTGTT 2117
QY 791 TTC-----TTGTGTTGTTCTTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTT 841
Db 2118 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2177
QY 842 TGAAGAGAGCAGTAGATGTTCTCTGTCGGGTCTCTGAGGACCTGGGAGCTCAGGCT 901
Db 2178 TGAAGAGAGCAGTAGATGTTCTCTGTCGGGTCTCTGAGGACCTGGGAGCTCAGGCT 2237
QY 902 GGGATCTCCAAAGGAGTAGTGGCTATCAAAATCAAAAGTCCAGTTGTGGGGGAA 961
Db 2238 GGGATCTCCAAAGGAGTAGTGGCTATCAAAATCAAAAGTCCAGTTGTGGGGGAA 2297
QY 962 AACAAAGCAGGCCATTACCCAGAGAGCTGTCCGCTTCCCTCAGCCAGCTAGGCT 1021
Db 2298 AACAAAGCAGGCCATTACCCAGAGAGCTGTCCGCTTCCCTCAGCCAGCTAGGCT 2357
QY 1022 TTGAAGGAAACAAAGACAGCAAAATGATTGGGTCTCTGGGAGATTGAGCTAGA 1081
Db 2358 TTGAAGGAAACAAAGACAGCAAAATGATTGGGTCTCTGGGAGATTGAGCTAGA 2417
QY 1082 GCTCTCTCTCCCAATCCTCCCTCGGCTGAGGAACTAACAAAGGAAAAAAATG 1141
Db 2418 GCTCTCTCTCCCAATCCTCCCTCGGCTGAGGAACTAACAAAGGAAAAAAATG 2477
QY 1142 CGAAGAGAGGATTAGAGAGAGCAAAATCCACTGTGTGCTTCCCTCAGCCAGCTAGG 1201
Db 2478 CGAAGAGAGGATTAGAGAGAGCAAAATCCACTGTGTGCTTCCCTCAGCCAGCTAGG 2537
QY 1202 ACTAGAGAGTCTGCGGAGCCGAGCCAGCGCTTCCGCGCTCTAGSCCGGCG 1261
Db 2538 ACTAGAGAGTCTGCGGAGCCGAGCCAGCGCTTCCGCGCTCTAGSCCGGCG 2597
QY 1262 GCCCGCGCGGGAAGGGACGACAGCCGCGGACCTTAAGACACCTGCTTACCCCTCAC 1321
Db 2598 GCCCGCGCGGGAAGGGACGACAGCCGCGGACCTTAAGACACCTGCTTACCCCTCAC 2657
QY 1322 GCCACCCACACG-----ACCTGCCGCCCACTCCCTAGATGTGTGCTGGCGGCTGAACG 1376

```

```

Db 2658 CCCACCCCCACCCACCCACCTCCCCCACTCCCTAGATGTGTGCTGGGGGTGAAAG 2717
QY 1377 TCGCCCGTTTAAAGGGCGGGCCCGCTCCACGTGCTTTCTGCTGAGTGACTGAATACA 1436
Db 2718 TCGCCCGTTTAAAGGGCGGGCCCGCTCCACGTGCTTTCTGCTGAGTGACTGAATACA 2777
QY 1437 TAAACAGAGCGCGGGAACGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGTAA 1496
Db 2778 TAAACAGAGCGCGGGAAGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGTAA 2837
QY 1497 COTCTGCGCTCGGTCCAGCGCAATATATAAAGGAACAGTCCCGGCAAAACCCCGTAA 1556
Db 2838 COTCTGCGCTCGGTCCAGCGCAATCTAARAAGGAACAGTCCCGGCAAAACCCCGTAA 2897
QY 1557 TTGCGAGCGAGAG 1569
Db 2898 TTGCGAGCGAGAG 2910

RESULT 5
ABL58400
ID ABL58400 standard; DNA: 1197 BP.
XX ABL58400;
AC
XX
XX
DT 30-JUL-2002 (first entry)
XX
DE Human large ATP-binding cassette transporter 1(hABC1) promoter sequence.
XX
XX Human; large ATP-binding cassette transporter 1; ABC1; promoter;
XX anti-arteriosclerotic; gene transfer; transactivator; ds.
XX
XX Homo sapiens.
XX
XX WO200183506-A1.
XX
XX 08-NOV-2001.
XX
XX 27-APR-2001: 2001WO-US-3654.
XX
XX 28-APR-2000: 2000US-0560372.
XX
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX
XX Tall AR:
XX
XX WPI: 2002-049334/06.
XX
XX Novel isolated human large ATP-binding cassette transporter 1 promoter
XX capable of directing transcription of heterologous coding sequence
XX positioned downstream to it, useful for expressing foreign DNA in host
XX cells.
XX
XX Claim 1; Fig 3: 68pp; English.
XX
XX The invention relates to an isolated human large ATP-binding cassette
XX transporter 1 (Abc1) promoter capable of directing transcription of
XX heterologous coding sequence positioned downstream to it. The Abc1
XX promoter is useful for expressing foreign DNA in a host cell, by
XX introducing into the host cell a gene transfer vector comprising the
XX promoter operably linked to a foreign DNA encoding a desired polypeptide
XX or RNA, where the foreign DNA is expressed. The gene transfer can be
XX introduced into the host cell by adenovirus infection, liposome-mediated
XX transfer, topical application to the cell or microinjection. The gene
XX transfer vector encodes and expresses a reporter molecule. The method
XX further involves introducing into the cell a gene transfer vector
XX comprising a nucleic acid segment encoding a transactivator protein
XX capable of upregulating the Abc1 promoter, or contacting the cell with
XX the transactivator protein, or an agonist of the transactivator protein.
XX Modulators of human Abc1 gene expression are useful for treating
XX atherosclerosis. The present sequence represents the hABC1 promoter.
XX
XX Sequence 1197 BP; 284 A; 314 C; 328 G; 271 T; 0 other;

```

Query Match									
Best Local Similarity 59.8%; Score 1146.4; DB 24; Length 1197;									
Matches 1180; Conservative 0; Mismatches 1; Indels 3; Gaps 3;									
QY	460	ACCTGAGTTTGGCCAGATAAGCTACATTTAGTTTGGTTGATGATGACATTA	519						
DB	1	ACCTGAGTTTGGCCAGATAAGCTACATTTAGTTTGGTTGATGATGACATTA	60						
QY	520	TATTTAGACATATGGTGTGAGCCCTGCATCTACTCTTTCCTTTTTCCTTCC	579						
DB	61	TAATTTAGACATATGGTGTGAGCCCTGCATCTACTCTTTCCTTTTTCCTTCC	120						
QY	580	AGTGTCTTGGGTAGTTTGTGTCCTCCCTACAGCCAAAGGCGAAGATGAGTGT	639						
DB	121	AGTGTCTTGGGTAGTTTGTGTCCTCCCTACAGCCAAAGGCGAAGATGAGTGT	179						
QY	640	GGAGTGGCTACATAATTTACAGACGCTGCAATCTCTGGCTGACCTTCAAAAT	699						
DB	180	GGAGTGGCTACATAATTTACAGACGCTGCAATCTCTGGCTGACCTTCAAAAT	239						
QY	700	AACTAATAATACAGTCTCTGTTTATACAGGAGGCTGATCAATATAATGAAT	759						
DB	240	AACTAATAATACAGTCTCTGTTTATACAGGAGGCTGATCAATATAATGAAT	299						
QY	760	AAGGGGCTGGTCCCATATGTTCTGTGTTTCTGTTTCTTCTTCTTCTTCTT	819						
DB	300	AAGGGGCTGGTCCCATATGTTCTGTGTTTCTGTTTCTTCTTCTTCTTCTT	358						
QY	820	TGGCTCTCTCTCTCAATTTATGAAGAGAGCAGTAGATGTTCTCTCGGTCCT	879						
DB	359	TGGCTCTCTCTCTCAATTTATGAAGAGAGCAGTAGATGTTCTCTCGGTCCT	418						
QY	880	AGGACCTGGGAGCTAGGCTGGGAATCTCAAGAGCAGTAGTGGCTATCAAAAT	939						
DB	419	AGGACCTGGGAGCTAGGCTGGGAATCTCAAGAGCAGTAGTGGCTATCAAAAT	478						
QY	940	AAGTCCAGGTTTGGGGGGGAAACAAAGAGCCCATATCCAGAGGACTTCCGCT	999						
DB	479	AAGTCCAGGTTTGGGGGGGAAACAAAGAGCCCATATCCAGAGGACTTCCGCT	538						
QY	1000	CCCTCACCCAGCTAGGCTTTGAAGAGAAACAAAGACAGCAAAATGATGGCT	1059						
DB	539	CCCTCACCCAGCTAGGCTTTGAAGAGAAACAAAGACAGCAAAATGATGGCT	598						
QY	1060	CCTGAGGAGATTACGCTTAGACTCTCTCTCCCAATTCCTCTCGGTGAGAA	1119						
DB	599	CCTGAGGAGATTACGCTTAGACTCTCTCTCCCAATTCCTCTCGGTGAGAA	557						
QY	1120	CTAACAGAGAAATAATTCGGNAACAGGATTTAGAGAGCAAAATTCACCTGGT	1179						
DB	658	CTAACAGAGAAATAATTCGGNAACAGGATTTAGAGAGCAAAATTCACCTGGT	717						
QY	1180	CCCTTGGCTGGGAGCTAGAGTCTGGGCGCAGCCCGAGCCGAGCGCTT	1239						
DB	718	CCCTTGGCTGGGAGCTAGAGTCTGGGCGCAGCCCGAGCCGAGCGCTT	777						
QY	1240	CCGCGCGCTTTAGCGCGCGGCGCGGCGGAGGAGCGAGCGCGGAGCCCTA	1299						
DB	778	CCGCGCGCTTTAGCGCGCGGCGCGGAGGAGCGAGCGCGGAGCCCTA	837						
QY	1300	AGACACTGCTGTACCTTCCACCCCAACCCACCTCCCTCCCTAGATGTG	1359						
DB	838	AGACACTGCTGTACCTTCCACCCCAACCCACCTCCCTCCCTAGATGTG	897						
QY	1360	TGCTGGCGGCTGAACCTGCGCTTTTAAAGGGCGGCGCGGCTGCTTTTCTGC	1419						
DB	898	TGCTGGCGGCTGAACCTGCGCTTTTAAAGGGCGGCGCGGCTGCTTTTCTGC	957						
QY	1420	TGAGTGAATGAACTACATAACAGAGCGCGGAGCGGAGGAGGAGGAGCACAG	1479						
DB	958	TGAGTGAATGAACTACATAACAGAGCGCGGAGCGGAGGAGGAGGAGCACAG	1017						

Query Match 7.3%; Score 119.4; DB 21; Length 227;
Best Local Similarity 99.2%; Pred. No. 7.1e-23;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1523 ATAAAGGAAGTACTAGTCTCCGGCAAAACCCGCTAATTCGAGCGAGAGTGAAGTGGGCGG 1582
DB 1 ATAAAGGAAGTACTAGTCTCCGGCAAAACCCGCTAATTCGAGCGAGAGTGAAGTGGGCGG 60

```
Oy 1593 GACCGGAGAGCGGAGCGGACCCCTCTCTCCGGGCTCGCGCAGGGCAGGGCGGGAGCT 1642
      |||||
Db 61 GACCGGAGAGCGGAGCGGACCCCTCTCTCCGGGCTCGCGCAGGGCAGGGCGGGAGCT 120
Oy 1643 C 1643
Db 121 C 121

RESULT 8
AAD37267/C
ID AAD37267 standard; DNA; 2910 bp.
XX
AC AAD37267;
DT 21-AUG-2002 (first entry);
DE Human ABC1 transcription regulatory DNA #3.
XX
KW Human; ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
KW cholesterol metabolism; hypercholesterolaemia; antisense therapy; ds.
XX
OS Homo sapiens.
FH
FH Key
FH Location/Qualifiers
FH 12..23
FH /tag= a
FH /bound_moiety= "LMO2COM/MYOD"
FH 97..107
FH /tag= b
FH /bound_moiety= "DeltaEF1"
FH 110..125
FH /tag= c
FH /bound_moiety= "S3/NKX2.5"
FH 196..211
FH /tag= d
FH /bound_moiety= "S8"
FH 228..237
FH /tag= e
FH /bound_moiety= "GATA"
FH 399..410
FH /tag= f
FH /bound_moiety= "IK2"
FH 412..420
FH /tag= g
FH /bound_moiety= "LYF1"
FH 528..539
FH /tag= h
FH /bound_moiety= "LMO2COM/MYOD/DeltaEF1"
FH 549..556
FH /tag= i
FH /bound_moiety= "LYF1"
FH 558..568
FH /tag= j
FH /bound_moiety= "DeltaEF1"
FH 590..596
FH /tag= k
FH /bound_moiety= "LXX2.5"
FH 608..620
FH /tag= l
FH /bound_moiety= "NFY/CAAT"
FH 708..715
FH /tag= m
FH /bound_moiety= "MZFl"
FH 723..730
FH /tag= n
FH /bound_moiety= "MZFl"
FH 771..785
FH /tag= o
FH /bound_moiety= "HFH2/SRY/EV11"
FH 803..812
FH /tag= p
FH /bound_moiety= "CREBP1/VBP"
FH 831..837
FH /tag= r
FH /bound_moiety= "NKX2.5"
FH 1075..1089
FH /tag= s
FH /bound_moiety= "GATA"
FH 1173..1188
FH /tag= t
FH /bound_moiety= "LXRalpha/deltaEF1"
FH 1189..1209
FH /tag= u
FH /bound_moiety= "DeltaEF1/LYF1/IK2"
FH 1483..1491
FH /tag= v
FH /bound_moiety= "A24"
FH 1498..1514
FH /tag= w
FH /bound_moiety= "LMO2-COM/MYOD/deltaEF-/E47"
FH 1524..1545
FH /tag= x
FH /bound_moiety= "ZID/deltaEF1"
FH 1597..1607
FH /tag= y
FH /bound_moiety= "DeltaEF-"
FH 1622..1627
FH /tag= z
FH /bound_moiety= "PPAR"
FH 1632..1637
FH /tag= aa
FH /bound_moiety= "PPAR"
FH 1685..1698
FH /tag= ab
FH /bound_moiety= "USF/NMYC/MYCMAX"
FH 1787..1797
FH /tag= ac
FH /bound_moiety= "DeltaEF-"
FH 1809..1819
FH /tag= ad
FH /bound_moiety= "DeltaEF1"
FH 1822..1833
FH /tag= ae
FH /bound_moiety= "SRY"
FH 1840..1850
FH /tag= af
FH /bound_moiety= "AP1"
FH 1842..1856
FH /tag= ag
FH /bound_moiety= "HNF3beta"
FH 1978..1985
FH /tag= ah
FH /bound_moiety= "NKX2.5"
FH 2008..2016
FH /tag= ai
FH /bound_moiety= "PPAR/NKX2.5/PPAR"
FH 2019..2024
FH /tag= aj
FH /bound_moiety= "PPAR/NKX2.5/PPAR"
FH 2051..2059
FH /tag= ak
FH /bound_moiety= "GATA"
FH 2104..2111
FH /tag= al
FH /bound_moiety= "SOX5"
FH 2114..2152
FH /tag= am
FH /bound_moiety= "SYR/HFH/HNF3beta"
FH 2221..2228
FH /tag= an
FH /bound_moiety= "MZFl"
FH 2234..2249
FH /tag= ao
FH /bound_moiety= "IK2/NFkappaB/CREL"
FH 2259..2272
FH
```

FT	/*tag= ap	
FT	/bound_moiety=	"LMO2COM/GATA"
FT	2289..2306	
protein_bind		
FT	/*tag= ag	
FT	/bound_moiety=	"MZFL/SRY"
FT	2313..2318	
protein_bind		
FT	/*tag= ar	
FT	/bound_moiety=	"PPAR"
FT	2321..2326	
protein_bind		
FT	/*tag= as	
FT	/bound_moiety=	"PPAR"
FT	2335..2342	
protein_bind		
FT	/*tag= at	
FT	/bound_moiety=	"MZFL"
FT	2361..2384	
protein_bind		
FT	/*tag= au	
FT	/bound_moiety=	"HNF3beta/SRY/EV11"
FT	2426..2433	
protein_bind		
FT	/*tag= av	
FT	/bound_moiety=	"MZFL"
FT	2442..2451	
protein_bind		
FT	/*tag= aw	
FT	/bound_moiety=	"AP4"
FT	2455..2466	
protein_bind		
FT	/*tag= ax	
FT	/bound_moiety=	"SRY"
FT	2491..2498	
protein_bind		
FT	/*tag= ay	
FT	/bound_moiety=	"STAT"
FT	2524..2534	
protein_bind		
FT	/*tag= az	
FT	/bound_moiety=	"STAT/PPAR"
FT	2536..2541	
protein_bind		
FT	/*tag= ba	
FT	/bound_moiety=	"PPAR"
FT	2589..2600	
protein_bind		
FT	/*tag= bb	
FT	/bound_moiety=	"AP2"
FT	2610..2617	
protein_bind		
FT	/*tag= bc	
FT	/bound_moiety=	"MZFL"
FT	2634..2648	
protein_bind		
FT	/*tag= bd	
FT	/bound_moiety=	"LMO2COM/MYOD/E47"
FT	2657..2672	
protein_bind		
FT	/*tag= be	
FT	/bound_moiety=	"RREB1"
FT	2680..2698	
protein_bind		
FT	/*tag= bf	
FT	/bound_moiety=	"MZFL/CMYB"
FT	2728..2740	
protein_bind		
FT	/*tag= bg	
FT	/bound_moiety=	"SP1/GC"
FT	2743..2757	
protein_bind		
FT	/*tag= bh	
FT	/bound_moiety=	"USF/NMYC/ARNT"
FT	2758..2773	
protein_bind		
FT	/*tag= bi	
FT	/bound_moiety=	"NFE2A1"
FT	2774..2787	
protein_bind		
FT	/*tag= bj	
FT	/bound_moiety=	"XFD1/HFH"
FT	2794..2806	
protein_bind		
FT	/*tag= bk	
FT	/bound_moiety=	"GC/SP-1/MZF1"
FT		

```

Query Match          6.2%; Score 102.2; DB 24; Length 2910;
Best Local Similarity 65.9%; Pred. No. 1.5e-17;
Matches 164; Conservative 83; Mismatches 83; Indels 2; Gaps 1;

y      241  ATGCCACCTCATCTGCGCCAAACTCAGGTCAAATGTGAAGAGTCTAAATGTAATCTG 300
      1820  ATGTCACCTTATCTGGCCAAACATCAGTCAAATGTGAAGAGTCTAAATGTAATCTG 1761

```

301 CCCTCAAGGTGGCTACAAAGGATATCTTTGTCAAGGTAGGAGACCTTGTGGCOTTCACGT 360
||||| ||||| ||||| ||||| ||||| |||||
1760 CCTTGAGGATTCATAGAAGACTTCAGACAGACGCCGTAGAAGAGGCCAAGCAGGCCCTG 1701
361 GCATTCCAGGCGCTGGTTGGCCTCTTCCTACGGGTCTGTCTGACTTCTTATGAATCTC 420
1700 GAAGTGCACGTGGAGGC--CACAAAGTCTCTACCTTCACAAGATACTCTTGTAGCAC 1643
421 CCTTCAGGCGAGATTCATATATAGACTCTTCACAGTTTGACCTGAGTTTGGCCAGAATA 480
||||| ||||| ||||| ||||| ||||| |||||
1642 CTTGAAGGCGAGATTCACATTTAGACTCTTCACAGTTTGACCTGAGTTTGGCCAGAATG 1583
481 AGGTGACAT 489
||||| |||||
1582 AGTGGCAT 1574

RESULT 9
AAD37265/C
ID AAD37265 standard; DNA; 3231 BP.
XX AAD37265;
XX AC
XX CT
XX DE Human ABC1 transcription regulatory DNA #1.
XX XX
XX KW Human; ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
XX KW cholesterol metabolism; hypercholesterolaemia; antisense therapy; ds.
XX OS Homo sapiens.
XX WO200183745-A2.
XX 08-NOV-2001.
XX PD
XX PF 02-MAY-2001; 2001WO-EP05488.
XX PR 02-MAY-2000; 2000US-201280P.
XX PA (AVET) AVENTIS PHARMA SA.
XX PI Rosier-Montus M, Prades C, Lemoine C, Naudin L, Deneffe P;
XX PI Brewer B, Duverger N, Remaley A, Santamarina-Pozo S;
XX WPI; 2002-154404/20.

Isolated nucleic acid useful for modifying the ATP-binding cassette 1 (ABC1) and screening for candidate modulatory compounds or substances .

Claim 1; Page 130-131; 152pp; English.

The invention relates a nucleic acid which is capable of regulating the transcription of human ATP-binding cassette 1 (ABC1) gene, which is a casual gene for pathologies linked to a dysfunctioning of cholesterol metabolism, including diseases such as atherosclerosis. Polynucleotides of the invention are used to screen candidate molecules or substances that are capable of modulating the transcription of the ABC1 gene. The CC are used in antisense therapy. Compositions comprising sequences of the invention are used to treat hypercholesterolemia and atherosclerosis. The present sequence is human ABC1 transcription regulating DNA.

Sequence 3231 BP: 809 A; 773 C; 876 G; 773 T; 0 other;

Query Match	6.2%	Score 102.2	DB 24	Length 3231
Best Local Similarity	65.9%	Pred. No. 1.6e-17		
Matches 164	Conservative	0	Mismatches 83	Indels 2
Gaps 1				
241	ATGCCACGTCGATTCGTGCCAAAACCTCAGGTCAACAGTGTGAAGAGTCTAAATG	300		
1929	ATGCCACGTCGATTCGTGCCAAAACCTCAGGTCAACAGTGTGAAGAGTCTAAATG	300		

301 CCCTCAAGTGGCTCAAAAGGTAICTTTGTCAGGTAGGAGACCTTTGGCCCTCCACGT 360
 1760 CCCTGAAGGATTCATAGAGACICAGGACAGACCCGTAGAGAGGCCAAGCGCCCTG 1701
 361 GCATTCACAGGCGCTGCTGGCCCTCTCTAGGCGGTCTGCTCTGAGTCTCTATGAATCTC 420
 1700 GAAGTGCACGTGGAGGC--CACAAAGTCTCTCACTTGACAAAGATACCTTTGTAGCCAC 1643
 421 CCTTCAGGCGAGATCADATTAGACTCTTCACAGTTTGACTGAGTTTGGCCAGATA 480
 1642 CTTGAAGGCGAGATTCACATTAGACTCTTCACAGTTTGACTGAGTTTGGCCAGATA 1583
 481 AGGTGACAT 489
 1582 AGGTGCAT 1574

RESULT 10
 AAD21326
 ID AAD21326 standard; DNA; 7260 BP.
 XX
 AC AAD21326;
 XX
 DT 28-JAN-2002 (first entry)
 XX
 DE Human ATP binding cassette transporter 1 (ABCI) gene.
 XX
 KW Human: ATP binding cassette transporter 1; ABC1; coronary heart disease;
 KW dermatological; atherosclerosis; cardiovascular; inflammatory disease;
 KW psoriasis; lipid disorder; antibacterial; septic shock; gene therapy;
 KW immunosuppressive; lupus erythematosus; rheumatoid arthritis; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 321..7106
 FT FT /*tag= a
 FT FT /*product= "Human ABC1 protein"
 FT FT
 PN EP1136552-A1.
 XX
 PD 26-SEP-2001.
 XX
 PF 20-MAR-2000; 2000EP-0105820.
 XX
 PR 20-MAR-2000; 2000EP-0105820.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Schmitz G, Bodzioch M;
 XX
 DR WPI; 2001-640389/74.
 DR P-PSDB; AAE13022.
 XX
 PT New adenosine triphosphate binding cassette transporter-1 gene
 PT polymorphisms, useful for diagnosing and treating lipid disorders,
 PT cardiovascular diseases and inflammatory diseases
 XX
 PS Example 1; Fig 1; 48pp; English.
 XX
 CC The invention relates to four common polymorphisms in the gene encoding
 CC ATP-binding cassette transporter-1 (ABCI). ABC1 is associated with
 CC decreased ApoA-I mediated efflux of cholesterol. The polymorphisms in
 CC ABC1 directly affects cellular lipid homeostasis, which is a key factor
 CC in the atherogenic processes. The ABC1 polymorphisms are useful for
 CC diagnosing and treating lipid disorders, cardiovascular diseases
 CC (coronary heart disease, atherosclerosis) and inflammatory diseases
 CC (psoriasis, lupus erythematosus). The identification of ABC1 as a
 CC transporter for interleukin-beta (IL-beta) identifies this gene as
 CC a candidate for treatment of inflammatory diseases including rheumatoid
 CC arthritis and septic shock. The present sequence is human ABC1 gene.

SQ Sequence 7260 BP; 1834 A; 1755 C; 1905 G; 1756 T; 0 other;
 Query Match 6.0%; Score 98; DB 22; Length 7260;
 Best Local Similarity 100.0%; Pred. No. 3.4e-16;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1546 AAACCCCGCTAATTCGAGCGAGAGTGTAGTGGCGCGGACCCGAGAGCGGAGCCGACCC 1605
 DB 1 AAACCCCGCTAATTCGAGCGAGAGTGTAGTGGCGCGGACCCGAGAGCGGAGCCGACCC 60
 QY 1606 TTCTCTCCCGCGCTCCGCGACGCGGCGGCGGAGCTC 1643
 DB 61 TTCTCTCCCGCGCTCCGCGACGCGGCGGCGGAGCTC 98

RESULT 11
 AAI70315
 ID AAI70315 standard; cDNA; 7260 BP.
 XX
 AC AA-70315;
 XX
 DT 07-JAN-2002 (first entry)
 XX
 DE Human ATP binding cassette transporter 1 (ABCI) cDNA.
 XX
 KW ATP binding cassette transporter 1; ABC1; human; lipid disorder;
 KW cholesterol; cardiovascular disease; inflammatory disease;
 KW antiinflammatory; antilipemic; antipsoriatic; dermatological;
 KW Tangier disease; coronary heart disease; diagnosis; gene therapy;
 KW polymorphism; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 321..7106
 FT FT /*tag= a
 FT CDS 501..7106
 FT FT /*tag= b
 FT FT /*note= "alternative open reading frame of AAI70314"
 FT FT replace(976,A)
 FT FT /*tag= c
 FT FT variation replace(1516,C)
 FT FT /*tag= d
 FT FT variation replace(2969,G)
 FT FT /*tag= e
 FT FT variation replace(3836,C)
 FT FT /*tag= f
 XX
 PN EP1136554-A1.
 XX
 PD 26-SEP-2001.
 XX
 PF 24-MAR-2000; 2000EP-0105401.
 XX
 PR 24-MAR-2000; 2000EP-0106401.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Schmitz G, Bodzioch M;
 XX
 DR WPI; 2001-640389/74.
 DR P-PSDB; AAM50228.
 XX
 PT New adenosine triphosphate binding cassette transporter gene
 PT polymorphisms, useful for diagnosing and treating lipid disorders,
 PT cardiovascular diseases and inflammatory diseases
 XX
 PS Disclosure; Page 26-28; 41pp; English.
 XX
 CC The present sequence is that of cDNA encoding the human adenosine
 CC triphosphate (ATP) binding cassette transporter 1 (ABCI) protein
 CC (see AAM50227). The sequence includes an extended open reading
 CC frame (ORF) to that provided by the sequence in AAI70314, using

an alternative ATG codon as initiation codon and thereby adding an extra 40 N-terminal amino acids to the encoded ABC1 protein (see CC AAM50228). The invention provides 4 common polymorphisms in the CC ABC1 gene. These were identified by sequencing the ABC1 gene in CC different Tangier kindreds. In the variant genes (numbered as in CC AAI70314), G is changed to A at position 596. T is changed to C at CC position 1136. A is changed to G at position 2599 or G is changed CC to C at position 3456, or any combination of these. All of these CC polymorphisms alter the amino acid sequence of ABC1 and therefore CC may affect its function. The 2 most common polymorphisms (G596A) CC and A2599G) are both associated with a decreased in vitro ApoA-I CC mediated efflux of cholesterol from mononuclear phagocytes, a CC feature typical of Tangier disease. 3 of the variants (G596A, CC A2589G and G3456C) are significantly increased in a population of CC men having low high density lipoprotein-cholesterol levels and CC established coronary heart disease (CHD) relative to CHD-free CC control subjects. The use of the provided ABC1 polymorphisms for CC the diagnosis and treatment of lipid disorders, cardiovascular CC diseases, and inflammatory diseases (e.g. psoriasis, lupus CC erythematoses) is claimed. Modulation of ABC1 transcripts or CC proteins by antisense or ribozyme technology or RNA decoys is also CC claimed.

Query Match 6.0%; Score 98; DB 22; Length 7260;
Best Local Similarity 100.0%; Pred. No. 3.4e-16;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1546 AAACCCCGTAATTCGAGCGAGAGTGTAGTGGCGGGGACCCGAGCGGAGCGGACCC 1605
DB 1 AAACCCCGTAATTCGAGCGAGAGTGTAGTGGCGGGGACCCGAGCGGAGCGGACCC 50

QY 1606 TTCTCTCCGCGCTGCGCGAGCGAGCGGCGGAGGCTC 1643
DB 61 TTCTCTCCGCGCTGCGCGAGCGAGCGGCGGAGGCTC 98

RESULT 12
AAF24681/c
ID AAF24681 standard; DNA; 1643 BP.
XX AAF24681;
AC AAF24681;
D? 20-APR-2001 (first entry)
DZ Nucleotide sequence of the 5' flanking region of the human ABC1 gene.
KW Human; adenosine triphosphate binding cassette protein 1; ABC1;
KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
KW atherosclerosis; cholesterol transport; ss.
XX Homo sapiens.
XX
XX WO200078972-A2.
XX
XX 28-DEC-2000.
XX
XX 16-JUN-2000; 2000WO-US16765.
XX
XX 18-JUN-1999; 99US-0140264.
XX
XX 14-SEP-1999; 99US-0153872.
XX
XX 19-NOV-1999; 99US-0166573.
XX
XX (CVTH-); CV THERAPEUTICS INC.
XX
XX Lawn RM, Wade D, Garvin M;
PI WPI; 2001-137812/-4.
XX
XX Adenosine triphosphate (ATP) binding cassette (ABC) polynucleotide,
PT useful for the development of agents for the treatment of heart disease
PT

and other disorders associated with hypercholesterolemia and atherosclerosis.
Claim 1: Page 143-144; 215pp; English.
The present sequence represents the 5' flanking region of the human adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABC1 resides in cell membranes and utilizes ATP hydrolysis to transport a wide variety of substrates across the plasma membrane. ABC1 is a pivotal protein in the apolipoprotein-mediated mobilisation of intracellular cholesterol stores. ABC1 is defective in Tangier disease, a genetic disorder characterised by abnormal HDL-cholesterol metabolism. The ABC1 gene is localised to chromosome 9q22-9q31. The ABC1 genes and proteins are useful for developing pharmaceutical agents for the treatment of heart disease and other disorders associated with hypercholesterolemia. The ABC1 genes are useful for developing screening assays to screen for compounds that regulate the expression of genes associated with cholesterol transport. The genes and proteins are also useful for are also useful as diagnostic indicators of cardiovascular disease and other disorders associated with hypercholesterolemia.

Query Match 5.9%; Score 97; DB 22; Length 1643;
Best Local Similarity 61.8%; Pred. No. 3e-16;
Matches 154; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 241 ATGCCACCTCAATTCGGCCAAACTCAGGTCAAGTGTGAAGAGTCTAAATGTGAATCTG 300
DB 489 ATGTCACTTAATTCGGCCAAACTCAGGTCAAGTGTGAAGAGTCTAAATGTGAATCTG 430

QY 301 CCCTTCAAGTGGGTACAAAGGTATCTTTGTCAAGGTAGGAGACCTTTGGGCTCCACGT 360
DB 429 CCCTCAAGGAGATTCATAGAGACTCAGGACAGACCCGTAGAGAGGCCAAGAGGCC 370

QY 361 GCACCTTCAGGCGCTGTTGGCCCTCTTCTACGGGTCTCTCTGAGTCTTCTATGATCTC 420
DB 369 TGAAGAGTCAAGTGGAGGCCACAGGTCTCTTACCTTGACAAAGATACCTTTGTAGCCAC 310

QY 421 CCCTCAGGCGAGATTCATATTTAGACTCTTCAACAGTTTGACCTGAGTTTGGCCAGATA 480
DB 309 CTTGAAGGCGAGATTCACATTTAGACTCTTCAACAGTTTGACCTGAGTTTGGCCAGATG 250

QY 481 AGGTGACAT 489
DB 249 AGGTGGCAT 241

RESULT 13
AAF24703/c
ID AAF24703 standard; DNA; 1643 BP.
XX AAF24703;
AC AAF24703;
D? 20-APR-2001 (first entry)
DZ Nucleotide sequence of the 5' flanking region of the human ABC1 gene.
KW Human; adenosine triphosphate binding cassette protein 1; ABC1;
KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
KW atherosclerosis; cholesterol transport; ss.
XX Homo sapiens.
XX
XX WO200078971-A2.
XX
XX 28-DEC-2000.
XX
XX 16-JUN-2000; 2000WO-US16591.
XX
XX 18-JUN-1999; 99US-0140264.
XX
XX 14-SEP-1999; 99US-0153872.

PR 19-NOV-1999; 99US-0166573.
XX (CVTH-) CV THERAPEUTICS INC.
PA (UNIV) UNIV WASHINGTON.
XX
XX LAW: RM, Wade D, Oram JF, Garvin M;
XX WP: 2001-137811/14.
XX Adenosine triphosphate (ATP) binding cassette protein (ABC) 1
PT polynucleotides and polypeptides, useful for treatment of heart disease
PT and other disorders associated with hypercholesterolemia and
PT atherosclerosis
XX
XX Disclosure; Page 138-139; 21pp; English.
XX
XX The present sequence represents the 5' flanking region of the human
CC adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABC1
CC resides in cell membranes and utilizes ATP hydrolysis to transport a wide
CC variety of substrates across the plasma membrane. ABC1 is a pivotal
CC protein in the apolipoprotein-mediated mobilization of intracellular
CC cholesterol. ABC1 is defective in Tangier disease, a genetic
CC disorder characterized by abnormal HDL-cholesterol metabolism. The ABC1
CC gene is localised to chromosome 9q22-9q31. The ABC1 genes and proteins
CC are useful for developing pharmaceutical agents for the treatment of
CC heart disease and other disorders associated with hypercholesterolemia
CC and atherosclerosis. The genes are useful for developing screening assays
CC to screen for compounds that regulate the expression of genes associated
CC with cholesterol transport. The genes and proteins are also useful for
CC are also useful as diagnostic indicators of cardiovascular disease and
CC other disorders associated with hypercholesterolemia.
XX
SQ Sequence 1643 BP; 370 A; 433 C; 457 G; 403 T; 0 other;

Query Match 5.9%; Score 97; DB 22; Length 1643;
Best Local Similarity 61.8%; Pred. No. 3e-16; Mismatches 95; Indels 0; Gaps 0;
Matches 154; Conservative 0;
QY 241 ATGCCACCTCATCTCGCCAAACTCAGGTCAAACTGCAAGAGTCTAAATGTGAATCTG 300
DB 489 ATGTCACCTTATCTCGCCAAACTCAGGTCAAACTGCAAGAGTCTAAATGTGAATCTG 430
QY 301 CCTTCAAGGTGGCTACAAAGGTATCTTTGTCAAGTAGAGACCTTGTGGCCCTCCACGT 360
DB 429 CCTGAGGGGAGATTATAGAGACTTCAGGACAGACCCCTAGAGAGGCCAAGCAGGCC 370
QY 361 GCACCTCCAGGGCCTGCTTGGCTCTTCTACGGGTCTGCTCTGATCTCTTATGATCTC 420
DB 369 TGAAGTGCACGTGGAGGCCACAAAGGCTCTCTACCTTGACAAAGATACCTTGTAGCCAC 310
QY 421 CCTTACGGGCAGATTATTTAGACTCTTCACAGTTTGACCTGAGTTTGGCCCAATA 480
DB 309 CTTGAGGGGAGATTACATTAGACTCTTCACAGTTTGACCTGAGTTTGGCCCAATA 250
QY 481 AGGTGACAT 489
DB 249 AGGTGGCAT 241

RESULT 14
AAF92831/c
ID AAF92831 standard; DNA; 163999 BP.
XX
XX AAF92831;
XX
DT 17-MAY-2001 (first entry)
XX Human ABC1 genomic DNA.
XX
XX High density lipoprotein-cholesterol; HDL-C; cardiovascular; ABC1; ds.
XX Homo sapiens.
XX

PN WO200115676-A2.
XX
XX 06-MAR-2001.
XX
XX 01-SEP-2000; 2000WO-IB01492.
XX
XX 01-SEP-1999; 99US-0151977.
PR 15-MAR-2000; 2000US-0526193.
PR 23-JUN-2000; 2000US-0213958.
XX (UYER-) UNIV BRITISH COLUMBIA.
PA (XENO-) XENON GENETICS INC.
XX
XX Hayden MR, Brooks-Wilson AR, Pimstone SN, Clee SM;
XX WP: 2001-244356/25.
XX
XX Treating a lower than normal high density lipoprotein-cholesterol
PT (HDL-C) level, a higher than normal triglyceride level, or a
PT cardiovascular disease, by administering a compound that modulates LXR-
PT or RXR-mediated transcriptional activity
XX
XX Claim 8; Fig 1; 317pp; English.
XX
XX The present invention relates to a method for treating a patient
CC diagnosed as having a lower than normal high density
CC lipoprotein-cholesterol (HDL-C) level, a higher than normal
CC triglyceride level, or a cardiovascular disease, involving
CC administering a compound that modulates LXR- or RXR-mediated
CC transcriptional activity or ABC1 expression or activity.
CC The LXR gene product may be used in an assay to identify
CC compounds useful for the treatment of a disease or condition selected a
CC lower than normal HDL cholesterol level, a higher than normal
CC triglyceride level, and a cardiovascular disease.
XX
XX Sequence 183999 BP; 49549 A; 37944 C; 41170 G; 54950 T; 386 other;
SQ
Query Match 5.6%; Score 92.6; DB 22; Length 183999;
Best Local Similarity 63.1%; Pred. No. 5.3e-14;
Matches 159; Conservative 0; Mismatches 90; Indels 3; Gaps 1;
QY 241 ATGCCACCTCATCTCGCCAAACTCAGGTCAAACTGCAAGAGTCTAAATGTGAATCTG 300
DB 27714 ATGTCACCTTATCTCGCCAAACTCAGGTCAAACTGCAAGAGTCTAAATGTGAATCTG 27655
QY 301 CCTTCAAGGTGGCTACAAAGGTATCTTTGTCAAGTAGAGACCTTGTGGCCCTCCACGT 360
DB 27654 CCTGAGGGGAGATTATAGAGACTTCAGGACAGACCCCTAGAGAGGCCCAAGCAGG 27595
QY 361 GCACCTCCAGGGCCTGCTTGGCTCTTCTACGGGTCTGCTCTGATCTCTTATGAT 417
DB 27594 CCTGGAAGTGCACGTGGAGGCCACAAAGGCTCTCTACCTTGACAAAGATACCTTGTAGC 477
QY 418 CTCCCTTACGGGCAGATTATTTAGACTCTTCACAGTTTGACCTGAGTTTGGCCCA 477
DB 27534 CACCTTGAGGGCAGATTACATTAGACTCTTCACAGTTTGACCTGAGTTTGGCCCA 27475
QY 478 ATAGGTGACAT 489
DB 27474 ATGAGGTGGCAT 27463

RESULT 15
AAD37268
ID AAD37268 standard; DNA; 221 BP.
XX
XX AAD37268;
XX
XX 22-AUG-2002 (first entry)
DT Human ABC1 gene exon 1A.
XX Human: ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
XX

KW cholesterol metabolism; hypercholesterolaemia; antisense therapy;
KW exon 1A; ds.

XX Homo sapiens.

XX WO200183746-A2.

XX PD

XX 08-NOV-2001.

XX 02-MAY-2001; 23C1WO-EP05488.

XX 02-MAY-2000; 2000US-201280P.

XX (AVET) AVENTIS PHARMA SA.

XX Rosier-Ventus M, Prades C, Lemoine C, Naudin L, Deneffe P;

PI Brewer B, Duverger N, Remaley A, Santamarina-Pojo S;

XX WPI; 2002-154404/20.

PT Isolated nucleic acid useful for modifying the ATP-binding cassette 1
PT (ABCL) and screening for candidate modulatory compounds or substances

XX Claim 4; Page 132; 152pp; English.

XX The invention relates a nucleic acid which is capable of regulating the
CC transcription of human ATP-binding cassette 1 (ABCL) gene, which is a
CC causal gene for pathologies linked to a dysfunctioning of cholesterol
CC metabolism, including diseases such as atherosclerosis. Polynucleotides
CC of the invention are used to screen candidate molecules or substances
CC that are capable of modulating the transcription of the ABCL gene. They
CC are used in antisense therapy. Compositions comprising sequences of the
CC invention are used to treat hypercholesterolaemia and atherosclerosis.
XX The present sequence is human ABCL gene exon 1A.

SQ Sequence 221 BP; 44 A; 62 C; 73 G; 42 T; 0 other;

Query Match 5.5%; Score 91; DB 24; Length 221;
Best Local Similarity 100.0%; Pred. No. 5.1e-15;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1553 GTAATGGCAGCGAGAGTGAGTGGGGGGGGGACCGCGAGAGCGCCGACCTCTCTC 1612

DB 1 GTANTGGAGCGAGAGTGAGTGGGGGGGGGACCGCGAGAGCGCCGACCTCTCTC 60

QY 1613 CCGGGCTGGGCGAGGGCGAGGGGGGGGAGCTC 1643

DB 61 CCGGGCTGGGCGAGGGCGAGGGGGGGGAGCTC 91

Search completed: February 3, 2003, 16:20:03
Job time : 972.868 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 12:56:40 ; Search time 16.5211 Seconds
(without alignments)
6534.066 Million cell updates/sec

Title: US-09-596-141C-3_COPY_1292_1643

Perfect score: 352

Sequence:

1 gacacctaacacactctg.....cagggcagggcgaggagctc 352

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCFUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	35	9.9	7218	1	US-08-232-463-14
C 2	34.2	9.7	1610	4	US-09-276-331-56
C 3	32.6	9.3	3250	4	US-06-122-126B-14
C 4	32.2	9.1	16389	4	US-09-741-154-3
C 5	31.6	9.0	1835	4	US-09-485-549-1
C 6	31.4	8.9	1817	2	US-08-870-518-8
C 7	31.4	8.9	2060	3	US-09-205-143-1
C 8	31.4	8.9	81001	4	US-09-750-580-1
C 9	31.2	8.9	975	1	US-08-671-325B-9
C 10	31.2	8.9	975	1	US-08-672-109B-9
C 11	31.2	8.9	975	1	US-08-842-045-9
C 12	31.2	8.9	975	1	US-08-842-238-9
C 13	31.2	8.9	975	3	US-08-629-335B-9
C 14	31.2	8.9	978	3	US-08-706-261A-17
C 15	31.2	8.9	978	4	US-09-097-231-17
C 16	31.2	8.9	441529	4	US-09-103-840A-1
C 17	30.6	8.7	1969	4	US-09-038-628-1
C 18	30.6	8.7	17425	4	US-09-511-625B-5
C 19	30.6	8.7	43280	2	US-08-804-227C-1
C 20	30.2	8.6	1154	3	US-08-651-136C-7
C 21	30.2	8.6	1154	4	US-09-228-911A-7
C 22	30	8.5	966	3	US-08-469-318-68
C 23	30	8.5	966	3	US-08-469-318-70
C 24	30	8.5	966	3	US-08-469-318-71
C 25	30	8.5	966	3	US-08-468-609A-68
C 26	30	8.5	966	3	US-08-468-609A-70
C 27	30	8.5	966	3	US-08-468-609A-71

C 28	30	8.5	966	4	US-08-446-872A-68	Sequence 68, Appl
C 29	30	8.5	966	4	US-08-446-872A-70	Sequence 70, Appl
C 30	30	8.5	966	4	US-08-446-872A-71	Sequence 71, Appl
C 31	30	8.5	966	4	US-08-762-227A-68	Sequence 68, Appl
C 32	30	8.5	966	4	US-08-762-227A-70	Sequence 70, Appl
C 33	30	8.5	966	4	US-08-762-227A-71	Sequence 71, Appl
C 34	30	8.5	966	5	PCT-US95-01185-68	Sequence 68, Appl
C 35	30	8.5	966	5	PCT-US95-01185-70	Sequence 70, Appl
C 36	30	8.5	966	5	PCT-US95-01185-71	Sequence 71, Appl
C 37	30	8.5	35100	1	US-08-306-831B-19	Sequence 19, Appl
C 38	30	8.5	35100	5	PCT-US93-06251-19	Sequence 19, Appl
C 39	29.8	8.5	8310	3	US-06-870-126-11	Sequence 11, Appl
C 40	29.8	8.5	8310	4	US-09-445-247-11	Sequence 11, Appl
C 41	29.8	8.5	14985	-	US-08-652-972A-6	Sequence 6, Appl
C 42	29.8	8.5	14985	5	PCI-US86-06231A-6	Sequence 6, Appl
C 43	29.6	8.4	2224	3	US-09-109-204-6	Sequence 6, Appl
C 44	29.6	8.4	68750	3	US-09-335-409-1	Sequence 1, Appl
C 45	29.6	8.4	68750	4	US-09-568-102-1	Sequence 1, Appl

ALIGNMENTS

RESULTS 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1900 Diagonal Road, Suite 500
City: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BERT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)683-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: p72gpt-F1s
US-08-232-463-14

Query Match 9.9% Score 35; DB 1; Length 7218;

Qy 121 C 121

Db 1299 C 1299

RESULT 8

```

US-09-750-560-1/C
: Sequence 1, Application US/09750580
: Patent No. 6455280
: GENERAL INFORMATION:
: APPLICANT: Yen, Frances
: APPLICANT: Denison, Blake
: APPLICANT: Bour, Barbara
: APPLICANT: Bihsain, Bernard
: APPLICANT: Dumas Milne Edwards, Jean-Baptiste
: APPLICANT: Duclert, Aymeric
: APPLICANT: Bouguerelet, Lydie
: APPLICANT: Ebbets-Reed, Dana
: APPLICANT: Salter-Cid, Luisa
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH
: FILE REFERENCE: 89 US2.CIP
: CURRENT APPLICATION NUMBER: US/09/750,560
: CURRENT FILING DATE: 2000-12-28
: PRIOR APPLICATION NUMBER: US 09/599,362
: PRIOR FILING DATE: 2000-06-21
: PRIOR APPLICATION NUMBER: PCT/IB00/0101
: PRIOR FILING DATE: 2003-06-21
: PRIOR APPLICATION NUMBER: PCT/IB99/02058
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: US 49/469/099
: PRIOR FILING DATE: 1999-12-21
: PRIOR APPLICATION NUMBER: US 60/113,686
: PRIOR FILING DATE: 1998-12-22
: PRIOR APPLICATION NUMBER: US 60/141,032
: PRIOR FILING DATE: 1999-06-25
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: Patent.pm
: SEQ ID NO 1
: LENGTH: 81001
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 10946..13946
: OTHER INFORMATION: 5'regulatory region
: NAME/KEY: exon
: LOCATION: 12947..12958
: OTHER INFORMATION: exon 1
: NAME/KEY: exon
: LOCATION: 13470..13526
: OTHER INFORMATION: exon 2
: NAME/KEY: exon
: LOCATION: 13641..13752
: OTHER INFORMATION: exon 3
: NAME/KEY: exon
: LOCATION: 14271..15968
: OTHER INFORMATION: exon 4
: NAME/KEY: misc_feature
: LOCATION: 15969..17969
: OTHER INFORMATION: 3'regulatory region
: NAME/KEY: allele
: LOCATION: 1239
: OTHER INFORMATION: 20-828-311 : polymorphic base C or T
: NAME/KEY: allele
: LOCATION: 12347
: OTHER INFORMATION: 17-42-319 : polymorphic base C or T
: NAME/KEY: allele
: LOCATION: 15241
: OTHER INFORMATION: 17-41-250 : polymorphic base C or T
: NAME/KEY: allele
: LOCATION: 42218
: OTHER INFORMATION: 20-841-149 : polymorphic base A or G
: NAME/KEY: allele

```

```

: LOCATION: 45442
: OTHER INFORMATION: 20-842-115 : polymorphic base A or G
: NAME/KEY: allele
: LOCATION: 77058
: OTHER INFORMATION: 20-853-415 : polymorphic base C or T
: NAME/KEY: primer_bind
: LOCATION: 929..949
: OTHER INFORMATION: 20-828.pu
: NAME/KEY: primer_bind
: LOCATION: 1357..1377
: OTHER INFORMATION: 20-828.rp complement
: NAME/KEY: primer_bind
: LOCATION: 12029..12050
: OTHER INFORMATION: 17-42.pu
: NAME/KEY: primer_bind
: LOCATION: 12581..12603
: OTHER INFORMATION: 17-42.rp complement
: NAME/KEY: primer_bind
: LOCATION: 14992..15012
: OTHER INFORMATION: 17-41.pu
: NAME/KEY: primer_bind
: LOCATION: 15480..15482
: OTHER INFORMATION: 17-41.rp complement
: NAME/KEY: primer_bind
: LOCATION: 42070..42090
: OTHER INFORMATION: 20-841.pu
: NAME/KEY: primer_bind
: LOCATION: 42572..42591
: OTHER INFORMATION: 20-841.rp complement
: NAME/KEY: primer_bind
: LOCATION: 45328..45347
: OTHER INFORMATION: 20-842.pu
: NAME/KEY: primer_bind
: LOCATION: 45863..45883
: OTHER INFORMATION: 20-842.rp complement
: NAME/KEY: primer_bind
: LOCATION: 76644..76664
: OTHER INFORMATION: 20-853.pu
: NAME/KEY: primer_bind
: LOCATION: 77166..77185
: OTHER INFORMATION: 20-853.rp complement
: NAME/KEY: primer_bind
: LOCATION: 1220..1238
: OTHER INFORMATION: 20-828-311.mis
: NAME/KEY: primer_bind
: LOCATION: 1240..1258
: OTHER INFORMATION: 20-828-311.mis complement
: NAME/KEY: primer_bind
: LOCATION: 12326..12346
: OTHER INFORMATION: 17-42-319.mis
: NAME/KEY: primer_bind
: LOCATION: 12348..12366
: OTHER INFORMATION: 17-42-319.mis complement
: NAME/KEY: primer_bind
: LOCATION: 15222..15240
: OTHER INFORMATION: 17-41-250.mis
: NAME/KEY: primer_bind
: LOCATION: 15242..15260
: OTHER INFORMATION: 17-41-250.mis complement
: NAME/KEY: primer_bind
: LOCATION: 42199..42217
: OTHER INFORMATION: 20-841-149.mis
: NAME/KEY: primer_bind
: LOCATION: 42219..42237
: OTHER INFORMATION: 20-841-149.mis complement
: NAME/KEY: primer_bind
: LOCATION: 45423..45441
: OTHER INFORMATION: 20-842-115.mis
: NAME/KEY: primer_bind
: LOCATION: 45443..45461
: OTHER INFORMATION: 20-842-115.mis complement
: NAME/KEY: primer_bind
: LOCATION: 77039..77057

```



```
Db 1359 CAGGTGGGGCCGTGGTCCACAGCAGGTGATCTCGTCCAGGTAGCGGGGACCGCAGTC 1300
QY 121 C 121
Db 1299 C 1299

RESULT 8
US-09-764-868-142
; Sequence 142, Application US/09764866
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 142
; LENGTH: 2928
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (491)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-142

Query Match 8.9%; Score 31.4; DB 9; Length 2928;
Best Local Similarity 56.2%; Pred. No. 6;
Matches 59; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1 GGACCTTAGACACCTGCTGTACCTCCACCCGCCACCCACCCACCTCCGCCCACTCC 60
Db 2464 GTACACTAGATCTCTGCTGGCCGCTCCAGGCTGGACCGACCCGCCACCCCTCT 2523
QY 61 TAGATGTGCTGGGGGGTGAACGTCGCCGCTTTAAGGGGGGG 105
Db 2524 GCTCTGATCCGTGGGCTCAGCAGCCATGCCCTGGGTGACGG 2568

RESULT 9
US-09-751-877-1/C
; Sequence 1, Application US/09751877
; Patent No. US20020142949A1
; GENERAL INFORMATION:
; APPLICANT: Yen, Frances
; APPLICANT: Denison, Blake
; APPLICANT: Bout, Barbara
; APPLICANT: Bihain, Bernard
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bouqueleret, Lydie
; APPLICANT: Ebbets-Reed, Dana
; APPLICANT: Salter-Cid, Luisa
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH
; FILE REFERENCE: 89.US3.REG
; CURRENT APPLICATION NUMBER: US/09/751,877
; CURRENT FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 8-001
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 10946..12946
; OTHER INFORMATION: 5'-regulatory region
; NAME/KEY: exon
; LOCATION: 12947..12958
; OTHER INFORMATION: 20-828-311.mis complement

; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 13470..13526
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 13641..13752
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 14271..15968
; OTHER INFORMATION: exon 4
; NAME/KEY: misc.feature
; LOCATION: 15969..17969
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 1239
; OTHER INFORMATION: 20-828-311 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 12347
; OTHER INFORMATION: 17-42-319 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 15241
; OTHER INFORMATION: 17-41-250 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 42218
; OTHER INFORMATION: 20-841-149 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 45442
; OTHER INFORMATION: 20-842-115 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 77058
; OTHER INFORMATION: 20-853-4-5 : polymorphic base C or T
; NAME/KEY: primer_bind
; LOCATION: 929..949
; OTHER INFORMATION: 20-825.pu
; NAME/KEY: primer_bind
; LOCATION: 1357..1377
; OTHER INFORMATION: 20-828.rp complement
; NAME/KEY: primer_bind
; LOCATION: 12029..12050
; OTHER INFORMATION: 17-42.pu
; NAME/KEY: primer_bind
; LOCATION: 12581..12603
; OTHER INFORMATION: 17-42.rp complement
; NAME/KEY: primer_bind
; LOCATION: 14992..15012
; OTHER INFORMATION: 17-41.pu
; NAME/KEY: primer_bind
; LOCATION: 15460..15482
; OTHER INFORMATION: 17-41.rp complement
; NAME/KEY: primer_bind
; LOCATION: 42070..42090
; OTHER INFORMATION: 20-841.pu
; NAME/KEY: primer_bind
; LOCATION: 42572..42591
; OTHER INFORMATION: 20-841.rp complement
; NAME/KEY: primer_bind
; LOCATION: 45328..45347
; OTHER INFORMATION: 20-842.pu
; NAME/KEY: primer_bind
; LOCATION: 45863..45883
; OTHER INFORMATION: 20-842.rp complement
; NAME/KEY: primer_bind
; LOCATION: 76644..76664
; OTHER INFORMATION: 20-853.pu
; NAME/KEY: primer_bind
; LOCATION: 77166..77185
; OTHER INFORMATION: 20-853.rp complement
; NAME/KEY: primer_bind
; LOCATION: 1220..1238
; OTHER INFORMATION: 20-828-311.mis
; NAME/KEY: primer_bind
; LOCATION: 1240..1256
; OTHER INFORMATION: 20-828-311.mis complement
```



```

; LENGTH: 99916
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (-)...(99916)
; OTHER INFORMATION: C = A,T,C or G
US-09-816-095-3

Query Match      8.58; Score 30.4; DB 10; Length: 99916;
Best Local Similarity 49.48; Pred. No. 38;
Matches 79; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 17 GCTGTACCTCCACCCGCCACCCACCTCCGCCCAACTCCCTAGATGTTCGTGGGC 76
   || || || || || || || || || || || || || || || || || || ||
Db 3348 GCCGTCCTCTGCTGCACTCCGCCACCTCCGCTGCCGCTGGGGGTGAGGGGCTGTGTG 3407

QY 77 GCTGTACCTCCACCCGCCACCCACCTCCGCCCAACTCCCTAGATGTTCGTGGGC 136
   || || || || || || || || || || || || || || || || || || ||
Db 3408 GAGAGGATCACTGTGCTTAAGGGCGGGGAGTCTGCCCTGGGGCTTTCTGTGGGAGAT 3467

QY 137 TGAACATACATAAAGAGAGCGCGGGAACCGGGGGAGGA 176
   || || || || || || || || || || || || || || || || || || ||
Db 3468 TGTGTCAAGAGATAGACAGGCTGTGAGCGCGGGGATGA 3507

```

Search completed: February 3, 2003, 16:33:24
Job time : 152.868 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: February 3, 2003, 16:39:45 ; Search time 263.012 Seconds
(without alignments)
8559.210 Million cell updates/sec

Title: US-09-596-14lc-3_COPY_1394_1532

Perfect score: 139

Sequence: 1 gggcccggtccacgtgct.....agccgaatctataaaggaa 139

Scoring table: IDENTITY_NCC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estma:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pla:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	35	25.2	1863	17	GGA200110
2	33.2	23.9	257	17	TA376C03P
3	32.8	23.6	962	12	BG256511
4	32.4	23.3	712	13	BI890368
5	32.2	23.2	513	17	AQ209600
6	31.8	22.9	973	12	BG757993

SUMMARIES

RESULT 1

GGA200110

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

7 31.6 22.7 440 17 TA157G02P
8 31.6 22.7 516 17 AQ643297
9 31.6 22.7 742 13 BI193040
10 31.6 22.7 808 13 BI115921
11 31.2 22.6 782 17 CNS01YNS
12 31.2 22.4 437 17 BH103719
13 31.2 22.4 529 12 BG438512
14 31.2 22.4 875 13 BI133879
15 30.8 22.2 340 9 AA324852
16 30.8 22.2 846 17 AZ184185
17 30.8 22.2 1170 14 BM10841
18 30.6 22.0 886 9 AL547703
19 30.4 21.9 359 12 BG322077
20 30.4 21.9 872 12 BF254048
21 30 21.6 310 9 AL534701
22 30 21.6 528 12 BG081346
23 30 21.6 586 12 BF725798
24 30 21.6 611 13 BM011917
25 30 21.6 671 10 BE391123
26 30 21.6 677 10 BE272078
27 30 21.6 683 10 BE258315
28 30 21.6 693 10 SE389737
29 30 21.6 701 13 BI193161
30 30 21.6 705 14 BM991113
31 30 21.6 709 12 BG897930
32 30 21.6 718 13 BI561910
33 30 21.6 728 13 BI520520
34 30 21.6 771 9 AL556360
35 30 21.6 774 12 BG702611
36 30 21.6 827 12 BG821508
37 30 21.6 837 12 BG770188
38 30 21.6 849 14 BG651632
39 30 21.6 870 14 BG87942
40 30 21.6 892 12 BF312138
41 30 21.6 894 14 BG670036
42 30 21.6 901 14 BG431809
43 30 21.6 911 14 BG222206
44 30 21.6 917 12 BG31708
45 30 21.6 922 13 BM451890

ALIGNMENTS

GGA200110 1863 bp DNA linear GSS 08-JUL-2000
Gallus gallus anonymous sequence from Cosmid mapping to a
microchromosome (Cosmid 20 - Contig 12), genomic survey sequence.
AJ231825
GSS: genome survey sequence.
Chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus
1 (bases 1 to 1863)
Smith, J., Bruley, C.K., Paton, I.R., Dunn, I., Jones, C.T., Windsor, D.,
Morrice, D.R., Law, A.S., Masabanda, J., Sazanov, A., Waddington, D.,
Fries, R. and Burt, D.W.
Differences in gene density on chicken macrochromosomes and
microchromosomes
Anim. Genet. 31 (2), 96-103 (2000)
20244064
10782207
2 (bases 1 to 1863)
Smith, J., Bruley, C.K., Paton, I.R., Law, A.S., Masabanda, J.,
Waddington, D., Fries, R. and Burt, D.W.
Direct Submission
Submitted (12-APR-1998) Division of Molecular Biology, Roslin
Institute, Roslin, Midlothian EH25 9PS, Scotland, UK
Location/Qualifiers

```

source
1..1863
/organism="Gallus gallus"
/db_xref="taxon:9031"
/chromosome="microchromosome"
BASE COUNT      408 a   392 c   580 g   482 t
ORIGIN

Query Match      25.2%; Score 35; DB 17; Length 1863;
Best Local Similarity 55.3%; Pred. No. 4.5;
Matches 58; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 17 TGCCTTTCGTGAGTGAAGTACATACATAACAGAGCGCGGGGAGGAGG 76
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 239 TGATTACAGCAGTGAAGTGAAGTACATACATAGCAGCAGCAGTCTTGGGAGC 298

QY 77 GAGACACAGCGTTTACCGGTAGTACCTCTGCGCTCGTGCACCGCAATCTATAAAG 136
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 299 GAAGACAGAAATTCGAGCTCATTCGAGCCAGCATCCACCCCTGGAGGTGAATATTAAACAG 358

QY 137 GAA 139
      |||
DB 359 GAA 361

RESULT 2
TA376C03P/c
LOCUS      TA376C03P      257 bp      DNA      linear      GSS 13-DEC-2000
DEFINITION      T. brucei sheared genomic DNA clone 376c03, forward sequence,
                  genomic survey sequence.
ACCESSION      AL497552
VERSION        AL497552.1 GI:11873274
KEYWORDS       GSS.
SOURCE         Trypanosoma brucei.
ORGANISM       Trypanosoma brucei.
                Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
                Trypanosoma.
REFERENCE      1 (bases 1 to 257)
AUTHORS        Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
                Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
                Melville, S.E., Rajandream, M.A. and Barrell, B.G.
TITLE          Direct Submission
JOURNAL        Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
                project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
                Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
                nh@sanger.ac.uk
COMMENT        Constructed at the Institute for Genomic Research (TIGR),
                Rockville, MD. Genomic DNA isolated from a cloned population of
                Trypanosoma brucei (TREF927/4 GUNet 10.1) was mechanically sheared
                to give a tight size distribution (
                4 kb). The v = 1 method used for the library construction is
                described in detail in Smith, H. and Venter, J.C. (Making small
                insert libraries for whole genome shotgun sequencing projects. In
                Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
                Barrell, Oxford University Press, 1999).
                Email: nels@tigr.org
                Details of T. brucei sequencing at the Sanger Centre are available
                at http://www.sanger.ac.uk/projects/T_brucei/.
FEATURES       source
                Location/Qualifiers
                1..257
                /organism="Trypanosoma brucei"
                /strain="TREF927"
                /db_xref="taxon:5691"
                /clone="376c03"
BASE COUNT      74 a   72 c   54 g   57 t
ORIGIN

Query Match      23.9%; Score 33.2; DB 17; Length 257;
Best Local Similarity 57.8%; Pred. No. 9.9;
Matches 59; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 28 GAGTGAAGTGAATACATACAGCGCGGGAACCGCGGGGAGGAGGAGGACACAC 87
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 257 GGGTGCTGCACTACCGATGCTGTACGCGCGGAGTGAGAGGGGGGGAACGACTACTGC 198

```

```

QY 88 CTTTGACCGATAGTAACCTCTGGCTGGTGCAGCGGAATCT 129
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 197 GTTTGACCGCTCCTAACCACTAGTTTTCATTTGATGTTACCACTCT 156

RESULT 3
BG256511      962 bp      mRNA      linear      ESI 13-FEB-2001
LOCUS      602370051F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4478196 5',
DEFINITION      mRNA sequence.
ACCESSION      BG256511
VERSION        BG256511.1 GI:12766327
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 962)
AUTHORS        NIH-MGC http://mgc.nci.nih.gov/.
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
                Unpublished (1999)
JOURNAL        Contact: Robert Strausberg, Ph.D.
                Email: rstrauss@remail.nih.gov
                Tissue Procurement: ATCC
                cDNA Library Preparation: Life Technologies, Inc.
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLAM10308 row: e column: 13
                High quality sequence start: 2
                High quality sequence stop: 733.
FEATURES       source
                Location/Qualifiers
                1..962
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4478196"
                /clone_lib="NIH_MGC_92"
                /tissue_type="embryonal carcinoma, cell line"
                /db_host="DHI0B (phage-resistant)"
                /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                Average insert size 2.5 kb. Library enriched for
                full-length clones and constructed by Life Technologies.
                Note: this is a NIH_MGC Library."
BASE COUNT      193 a   276 c   314 g   178 t
ORIGIN

Query Match      23.6%; Score 32.8; DB 12; Length 962;
Best Local Similarity 67.6%; Pred. No. 18;
Matches 46; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 16 GTGCTTTCGTGCTGAGTGAAGTACATACATACAGAGCGCGGGGAGGAGGAG 75
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 780 GTACTTCCTGTGCGAGTTACAGGAACCGCTCCAGGGGGGCGCGGGGAGGAGTG 839

QY 76 GGAGAGCA 83
      ||| |||
DB 840 GGAGACACA 847

RESULT 4
BI890368/c
LOCUS      BI890368      712 bp      mRNA      linear      ESI 12-OCT-2001
DEFINITION      ZF637-2-002123 zebrafish shield stage whole embryo cDNA library
                MPMGP637 Danio rerio cDNA clone MPMGP637_2P1; MPMGP637P017 5', mRNA
                sequence.
ACCESSION      BI890368
VERSION        BI890368.1 GI:16097639
KEYWORDS       EST.
SOURCE         zebrafish.

```

```

Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3236 row: E column: 20
Class: BAC ends
High quality sequence stop: 513.
Location/Qualifiers
1..513
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=3236 Col=20 Row=E"
/clone_library="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in
E-Coli DH10B"
BASE COUNT 108 a 118 c 101 g 175 t 11 others
ORIGIN

Query Match 23.2%; Score 32.2; DB 17; Length 513;
Best Local Similarity 61.2%; Pred. No. 23;
Matches 52; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 36 GAACATCATACAGAGCCGGGACCGCGGGCGGAGGAGGAGAGCAGCGCTTTGACC 95
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 186 GAACACATACACATGTGGCGCTCTTGGGGGTGGGGGAGGAGGAGCATCAGGAAGATA 127
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 96 GATAGTAACTCTCGGTCGGTGCA 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 126 GATAAATAATCCTGGTTAGATACA 102
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

FEATURES
source
RESULT 6
BG757993/G

```

```

DEFINITION      mRNA sequence.
ACCESSION      BG757993
VERSION        BG757993.1  GI:14068646
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1  (bases 1 to 973)
AUTHORS       NIH-MGC http://mgc.nci.nih.gov/.
TITLE         National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT       Unpublished (1999)
               Contact: Robert Strausberg, Ph.D.
               Email: cgabbs-r@mail.nih.gov
               Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
               cDNA Library Preparation: Ling Hong/Rubin Laboratory
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: Incyte Genomics, Inc.
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: L1CM1704  row: 1  column: 09
               High quality sequence stop: 498.
               Location/Qualifiers
                 1..973
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone="IMAGE:4855376"
                   /clone_lib="NIH-MGC_48"
                   /tissue_type="primary B-cells from tonsils (cell line)"
                   /lab_host="DH10B (phage-resistant)"
                   /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                   Site_2: EcoRI; cDNA made by oligo-dT priming.
                   Directionally cloned into EcoRI/XhoI sites using the
                   following 5' adaptor: GGCACAG(5). Size-selected >500bp
                   for average insert size 1.8kb. Library constructed by Ling
                   Hong in the laboratory of Gerald M. Rubin (University of

```



```

ACCESSION BI193040
VERSION BI193040.1 GI:14649060
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 742)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CMI852 row: f column: 23
High quality sequence start: 15
High quality sequence stop: 393.
Location/Qualifiers
FEATURES
source
1..742
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5090326"
/tissue_type="epithelioid carcinoma cell line"
/ab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dt priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GCCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC Library."
BASE COUNT 175 a 240 c 197 g 130 t
ORIGIN

Query Match 22.7%; Score 31.6; DB 13; Length 742;
Best Local Similarity 62.8%; Pred. No. 38;
Matches 49; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 GGGCCCCGGCTCCACGTCCTTCGTGAGTGACTGACTACATAACAGAGCGCGGAA 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151 GGGCTGCTGCTCCGTCGTCCTCGGCGCTGATGGCGGACAGGTGATCAGGGGCGAGCGA 92
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 CGGGCGCGGGGAGGAGGA 78
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91 CGGGCGCGGGGAGGAGGA 74
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
BI115921/c
LOCUS BI115921 808 bp mRNA linear EST 26-JUN-2001
DEFINITION 602866693F1 NIH_MGC_7 Homo sapiens CDNA clone IMAGE:5015138 5',
mRNA sequence.
ACCESSION BI115921
VERSION BI115921.1 GI:14566822
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 808)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: c9apbs-remail.nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CMI823 row: j column: 23
High quality sequence stop: 500.
Location/Qualifiers
FEATURES
source
1..808
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5015138"
/tissue_type="small cell carcinoma"
/ab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GCCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 201 a 237 c 227 g 143 t
ORIGIN

Query Match 22.7%; Score 31.6; DB 13; Length 808;
Best Local Similarity 52.8%; Pred. No. 39;
Matches 49; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 1 GGGCCCCGGCTCCACGTCCTTCGTGAGTGACTGACTACATAACAGAGCGCGGAA 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150 GGGCTGCTGCTCCGTCGTCCTCGGCGCTGATGGCGGACAGGTGATCAGGGGCGAGCGG 91
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 CGGGCGCGGGGAGGAGGA 78
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 90 CGGGCGCGGGGAGGAGGA 73
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
CNS01YK5/c
LOCUS CNS01YK5 782 bp DNA linear GSS 12-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence I7 end of clone
218D02 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL173210
VERSION AL173210.1 GI:7811267
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 782)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 782)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 782)
AUTHORS Genoscope.

```

RPC1-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pPARS1 cloning vector at the BamHI sites using MboI partially digested male C57Bl/6J DNA. "

100 a	111 c	106 g	120 t
-------	-------	-------	-------

BASE COUNT
ORIGIN

[illegible]

QY 87 GCTTGACCGAATAGTA 102
|||
Db 141 TTCAATACGTAAAGTA 126

RESULT_13					
EG438512/c					
LOCUS		525 bp	mRNA	linear	EST "O-MAY-2601
DEFINITION					
ps27c04.y2 Trichinella spiralis CC1 CMySprot Jansner trichinella					
spiralis cDNA 5' similar to SW:CC4 DROME P40793 CC042 HOMOLOG.					[5]
IR:Q3V465 TR:Q9U956 TR:Q9U955 FR:Q9U954 ; mRNA sequence.					
EG438512					
ACCESSION					

RECORDS	SOURCE	ORGANISM	REFERENCE
1		<i>Trichinella spiralis</i> .	
1		<i>Trichinella spiralis</i>	
1		Eukaryota; Metazoa; Nematoda; Enopleia; Trichocephalida;	
1		Trichinellidae; <i>Trichinella</i> .	
1		(bases: 10, 52)	

AUTHORS
McCarter, J., Callilond, S., Chapell, B., Pope, D., Mullin, J., Wyler, I., Dante, M., Marra, M., Hillier, L., Kudach, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelshvili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, K., Kohn, S., Shint, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Willson, R.

TITLE	JOURNAL	COMMENT
The Washington Univ. Nematode EST Project, 1999	Unpublished (1999)	Contact: McCarter JP The Washington Univ. Nematode EST Project, 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA tel. 314 286 1800

ver. 314 280 1810
 fax: 314 286 1810
 Email: est@watson.wustl.edu
 The library was constructed by Dr. Doug Jasmer
 (djasmer@w-med.wsu.edu) at Washington State University, Dept. of
 Veterinary Microbiology and Pathology DNA Sequencing by: Washington
 University Genome Sequencing Center St. Louis.
 Seq primer: Sp6
 High quality sequence stop: 422.
 Location/Qualifiers
 1..529
 /organism="Trichinella spiralis"
 /db_xref="taxon:6334"
 /clone_lib="Trichinella spiralis ML Cmwspout jasmer"

```

/ab_host="DH08"
/note="Vector: pCMVSPORT-7.neo; Site_1: NotI; Site_2: SalI
; The library was constructed using mRNA isolated from
Total RNA with oligo-dT cellulose. Total RNA was generated
from muscle larvae that were isolated from infected rats.
Larvae were liberated by pepsin/HCl digestion, incubated
with 1% SDS, treated with RNase and DNase to eliminate
host nucleic acid contamination, and purified on a Percoll
gradient. The T. spiralis isolate was obtained from
Dickson Despotemper (Columbia University). The library was
provided by Dr. Doug Jasmer (djasmer@vetmed.wsu.edu) and

```

colleagues at Washington State University. DNA Sequencing
by: Washington University Genome Sequencing Center St.
Louis."

BASE COUNT 161 a 107 c 119 g 142 t
ORIGIN

Query Match 22.4%; Score 31.2; DB 13; Length 875;
Best Local Similarity 54.3%; Pred. No. 46;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 11 TCCACGCTTTCGCTGAGTACTAGTAAACAGAGCGCGGAGCGCGGG 70

DB 228 TTCACATTTTCATAGCTGCATGCGCAAACTACAGAAAACAGACGAGATCCGTC 169

QY 71 AGGAGGGAGACAGAGGCTTTGACGAGTAGTAACCTTCGCTCGGTCCAGCCGAA 126

DB 168 TGGGATAGGACAGTGGCTCAACAGCATCGTAATCTTCTGTCCGCGAGTGCAGAA 113

RESULT 14

BI193879

LOCUS 875 bp mRNA linear EST 10-JUL-2001
DEFINITION 502948207F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:5091625 5';
mRNA sequence.

ACCESSION BI193879

VERSION BI193879.1 GI:14648899

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 875)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cdapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNI)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/ILLNI at:

<http://image.llnl.gov>

Plate: LICM1855 row: m column: 02

High quality sequence stop: 738.

FEATURES

source

Location/Qualifiers

1..875

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5091625"

/clone_lib="NIH_MGC_42"

/tissue_type="epithelioid carcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;

Site_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(S). Size-selected >500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library. "

175 a 290 c 222 g 187 t

BASE COUNT 175 a 290 c 222 g 187 t

ORIGIN

Query Match 22.4%; Score 31.2; DB 13; Length 875;

Best Local Similarity 53.2%; Pred. No. 52;

Matches 66; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 65 GCGGGAGGAGGAGACACAGGCTTTGACCGATAGTAACCTTCGCGTCGGTGCAGCG 124

DB 141 GCAGAGGGGCGCCCGGAGGGGTGCTACTTCTTAAACACCTCTCGGGCTCTTAGTCA 200

QY 125 AATC 128

DB 201 CAGC 204

RESULT 15

AA324852

LOCUS 340 bp mRNA linear EST 20-APR-1997

DEFINITION EST27749 Cerebellum II Homo sapiens cDNA 5' end, mRNA sequence.

ACCESSION AA324852

VERSION AA324852.1 GI:1977096

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 340)

AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult

,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White

,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,

Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald

,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodek,A.,

Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., J., Kelley,J.M.,

Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,

Moreno-Palances,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,

Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,

Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,

Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,I.A., Collins,E.J.,

Dinkeld., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A., He,W.W.,

Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H., Raymond,L.,

Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon

,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and

Venter,J.C.

Initial assessment of human gene diversity and expression patterns

based upon 83 million nucleotides of cDNA sequence

Nature 377 (6547 Suppl), 3-174 (1995)

96026280

COMMENT

Contact: Kerlavage, AR

Bioinformatics

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA

Fax: 3018699056

Email: arkerlav@tigr.org

For clone availability, additional sequence and expression

information related to this EST, please check the TIGR Human Gene

Index (<http://www.tigr.org/tadb/hgi/hgi.html>)

Seq primer: M13 Reverse.

Location/Qualifiers

1..340

/organism="Homo sapiens"

/db_xref="ATCC (inhost):125343"

/db_xref="taxon:9606"

/clone_lib="Cerebellum II"

/tissue_type="cerebellum"

/dev_stage="adult"

/note="Organ: brain; Vector: pBluescript SK-; Site_1:

EcoRI; Site_2: XhoI"

57 a 129 c 70 t 6 others

BASE COUNT 57 a 129 c 70 t 6 others

ORIGIN

Query Match 22.2%; Score 30.8; DB 9; Length 340;

Best Local Similarity 52.1%; Pred. No. 55;

Matches 62; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 5 CCGGCTCCACGCTGCTTCTGCTGAGTCACTGAATACATAAACAAGCGCGGACGGG 64

DB 81 CCGGACCCACGCTCTGGACCCCGGA-TATTTACGCTTCAGACAGACGCGGCGGG 140

Db 163 CCCATNTCCGGAGAGGCCCGAGACCCCTGAACTACTTCTCGGTGGCTGGCGGGAG 222
QY 65 GCGGGAGAGGGAGAGCACAGGCTTTGACCGATAGTAACTCTGGGTGGGTGGAGCC 123
Db 223 GGTGRANTTGGGAGCAGCGCANAGTTGGCCCTGCCACCCCAACCTTCTCCCGAGCC 281

Search completed: February 4, 2003, 03:27:17
Job time : 277.726 secs


```
NAME/KEY: primer_bind
LOCATION: 77039..77057
OTHER INFORMATION: 20-853-415.mis
NAME/KEY: primer_bind
LOCATION: 77059..77077
OTHER INFORMATION: 20-853-415.mis complement
NAME/KEY: misc_binding
LOCATION: 227..1251
OTHER INFORMATION: 20-828-311.probe
NAME/KEY: misc_binding
LOCATION: 12335..12359
OTHER INFORMATION: 17-42-319.probe
NAME/KEY: misc_binding
LOCATION: 15229..15253
OTHER INFORMATION: 17-41-250.probe
NAME/KEY: misc_binding
LOCATION: 42206..42230
OTHER INFORMATION: 20-941-149.probe
NAME/KEY: misc_binding
LOCATION: 45430..45454
OTHER INFORMATION: 20-842-115.probe
NAME/KEY: misc_binding
LOCATION: 77046..77070
OTHER INFORMATION: 20-853-415.probe
US-09-750-580-1

Query Match      21.68; Score 30; DB 4; Length 8100;
Best Local Similarity 61.58; Pred. No. 1.9;
Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 1 GGGCCCGGGCCACGCTCTCTGCTGAGTGACTGAACTACATAAACAGAGCGCGGAA 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17479 GGGCTGCTGCTCTCTGCTGAGTGACTGAACTACATAAACAGAGCGCGGAA 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 61 GGGGGCGGGGAGGAGGGA 78
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17419 GGGGGCGGGGAGGAGGGA 17402

RESULT 3
US-09-276-531-56
Sequence 56, Application US/09276531
Patent No. 6183968
GENERAL INFORMATION:
APPLICANT: Bandmar, Olga
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Reddy, Roopa
APPLICANT: Guegler, Karl J.
APPLICANT: Baughn, Mariah R.
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING
RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION
NUMBER OF SEQUENCES: 134
CORRESPONDENCE ADDRESS:
ADDRESS: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/276.531
FILING DATE: Herewith
CLASSIFICATION:
PROR APPLICATION DATA:
APPLICATION NUMBER: 60/079,677
FILING DATE: March 27, 1998
```

```
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lynn E. Murty, Ph.D.
REGISTRATION NUMBER: 42,918
REFERENCE/DOCKET NUMBER: PA-0008 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 1610 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAITUT08
CLONE: 1396833
US-09-276-531-56

Query Match      20.44; Score 28.4; DB 4; Length 1610;
Best Local Similarity 53.68; Pred. No. 1.7;
Matches 59; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 2 GGGCCCGGCTCCACGTGCTTCTGCTGAGTGACTGAACTACATAAACAGAGCGCGGAAAC 61
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1211 GGCTGGGTATTCCTACTTCTGCTGAGTGACTGAACTACATAAACAGAGCGCGGAAAC 1270
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 62 GGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 111
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1271 CGAATGAGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1320
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
US-09-212-247C-3/c
Sequence 3, Application US/09212247C
Patent No. 6391603
GENERAL INFORMATION:
APPLICANT: POMPEJUS, Markus; SUTLBERGER, Harald; JOEFFKEN, Hans
and GARCIA, Maria Angeles Santos
and WOLFGANG, DOVAL, Jose Luis Revuelta; JIMENEZ, Alberto;
TITLE OF INVENTION: Genes of purine biosynthesis from Ashbya Gossypii
and the use thereof in microbial riboflavin
synthesis
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Keil & Weinkauff
STREET: 1101 Connecticut Avenue
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: IBM AT-compatible, Pentium processor
OPERATING SYSTEM: Windows 95
SOFTWARE: WordPerfect version 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/212,247C
FILING DATE: 16-Dec-1998
CLASSIFICATION: <Unknown>
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5369 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..54
FEATURE:
```



```

; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Multivariant IL-3 Hematopoiesis Fusion
; NUMBER OF SEQUENCES: 196
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,318
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/446,872
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 966 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-318-68

Query Match          19.3%; Score 26.8; DB 3; Length 966;
Best Local Similarity 50.0%; Pred. No. 5.1;
Matches 67; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 3 GCCCGGCTCCAGTCCTTCTGCTGAGTGACTGAACATACATAAAGAGAGCGGGAACG 62
DB 522 GCCCTGGATCTTCCTCACTTGCCTAAGCACTTGACGAGAACTCTGGGCGAGGAGCT 463

QY 63 GGGCGGGAGGAGGAGAGACAGAGCTTTGACGGATAGTAACCTCTGCCCTCGGTGCAGC 122
DB 462 GGCAGGGCCCTAAATGGTGTAGCAATGTTGGAGATTATGAGATCTTTAGACGGAGGAGA 403

QY 123 CGAATCTATAAAAG 136
DB 402 CGGGTTGATAGTAG 389

RESULT 9
US-08-469-318-71/c
; Sequence 71, Application US/08469318
; Patent No. 6022535
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Multivariant IL-3 Hematopoiesis Fusion
; NUMBER OF SEQUENCES: 196
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,318
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/446,872
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 966 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-318-71

Query Match          19.3%; Score 26.8; DB 3; Length 966;
Best Local Similarity 50.0%; Pred. No. 5.1;
Matches 67; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 3 GCCCGGCTCCAGTCCTTCTGCTGAGTGACTGAACATACATAAAGAGAGCGGGAACG 62
DB 522 GCCCTGGATCTTCCTCACTTGCCTAAGCACTTGACGAGAACTCTGGGCGAGGAGCT 463

QY 63 GGGCGGGAGGAGGAGAGACAGAGCTTTGACGGATAGTAACCTCTGCCCTCGGTGCAGC 122
DB 462 GGCAGGGCCCTAAATGGTGTAGCAATGTTGGAGATTATGAGATCTTTAGACGGAGGAGA 403

QY 123 CGAATCTATAAAAG 136
DB 402 CGGGTTGATAGTAG 389

RESULT 10
US-08-468-609A-68/c
; Sequence 68, Application US/08468609A
; Patent No. 6030812
; GENERAL INFORMATION:
; APPLICANT: Adams, Mark A.
; APPLICANT: Bauer, S. C.
; APPLICANT: Bradford-Goldberg, Sarah R.
; APPLICANT: Caparon, Maïre H.
```

```

; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Multivariant IL-3 Hematopoiesis Fusion
; NUMBER OF SEQUENCES: 196
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,318
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/446,872
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 966 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-469-318-71

Query Match          19.3%; Score 26.8; DB 3; Length 966;
Best Local Similarity 50.0%; Pred. No. 5.1;
Matches 67; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 3 GCCCGGCTCCAGTCCTTCTGCTGAGTGACTGAACATACATAAAGAGAGCGGGAACG 62
DB 522 GCCCTGGATCTTCCTCACTTGCCTAAGCACTTGACGAGAACTCTGGGCGAGGAGCT 463

QY 63 GGGCGGGAGGAGGAGAGACAGAGCTTTGACGGATAGTAACCTCTGCCCTCGGTGCAGC 122
DB 462 GGCAGGGCCCTAAATGGTGTAGCAATGTTGGAGATTATGAGATCTTTAGACGGAGGAGA 403

QY 123 CGAATCTATAAAAG 136
DB 402 CGGGTTGATAGTAG 389

RESULT 10
US-08-468-609A-68/c
; Sequence 68, Application US/08468609A
; Patent No. 6030812
; GENERAL INFORMATION:
; APPLICANT: Adams, Mark A.
; APPLICANT: Bauer, S. C.
; APPLICANT: Bradford-Goldberg, Sarah R.
; APPLICANT: Caparon, Maïre H.
```



```

; TITLE OF INVENTION: Fusion Proteins Comprising Multiply Mutated Interleukin-3 (IL-
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,
; STREET: P. O. Box 5110
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60680
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,609A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/192,325
; FILING DATE: 14-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Dennis A.
; REGISTRATION NUMBER: 34,547
; REFERENCE/DOCKET NUMBER: C-2790/3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)737-6986
; TELEFAX: (314)737-6972
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 966 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-468-609A-71

Query Match 19.3%; Score 26.8; DB 3; Length 966;
Best Local Similarity 50.0%; Pred. No. 5.1;
Matches 67; Conservative 0; Mismatches 67; Indels 0; Gaps 0:

QY 3 GCGCCGCTCCAGTGTCTCTGAGTGACCTGAACATAAAACAGAGCGCGGGAACG 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 522 GCGCTGAGTCTCTCTGAGTGACCTGAGGAGGAGCTCTGGGGCAGGAGCT 463

QY 63 GCGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 122
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 462 GCGAGGCGCTAATGGTGTAGCCATGTTGGAGATTATGAGATTCTTAGACGGAGGAGA 403

QY 123 CGAATCTATAAAG 136
   || ||| ||| |||
Db 402 CGGTTGATAGTAG 389

RESULT 13
US-08-446-872A-68/c
; Sequence 68, Application US/08446872A
; Patent No. 6361977
; GENERAL INFORMATION:
; APPLICANT: Abrams, Mark A.
; APPLICANT: Bauer, S. C.
; APPLICANT: Braford-Goldberg, Sarah R.
; APPLICANT: Caparon, Mairé H.
; APPLICANT: Easton, Alan M.
; APPLICANT: Klein, Barbara K.
; APPLICANT: McKearn, John P.
; APPLICANT: Orlins, Peter O.
; APPLICANT: Paik, Kumhan
; APPLICANT: Thomas, John W.
; TITLE OF INVENTION: Multivariant IL-3 Hematopoiesis
; TITLE OF INVENTION: Fusion Protein
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,
; STREET: P. O. Box 5110
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60680
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,872A
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/192,325
; FILING DATE: 14-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bennett, Dennis A.
; REGISTRATION NUMBER: 34,547
; REFERENCE/DOCKET NUMBER: C-2790/1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)737-6986
; TELEFAX: (314)737-6972
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 966 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: Gouble
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-446-872A-68

Query Match 19.3%; Score 26.8; DB 4; Length 966;
Best Local Similarity 50.0%; Pred. No. 5.1;
Matches 67; Conservative 0; Mismatches 67; Indels 0; Gaps 0:

QY 3 GCGCCGCTCCAGTGTCTCTGAGTGACCTGAACATAAAACAGAGCGCGGGAACG 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 522 GCGCTGAGTCTCTCTGAGTGACCTGAGGAGGAGCTCTGGGGCAGGAGCT 463

QY 63 GCGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 122
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 462 GCGAGGCGCTAATGGTGTAGCCATGTTGGAGATTATGAGATTCTTAGACGGAGGAGA 403

QY 123 CGAATCTATAAAG 136
   || ||| ||| |||
Db 402 CGGTTGATAGTAG 389

RESULT 14
US-08-446-872A-70/c
; Sequence 70, Application US/08446872A
; Patent No. 6361977
; GENERAL INFORMATION:
; APPLICANT: Abrams, Mark A.
; APPLICANT: Bauer, S. C.
; APPLICANT: Braford-Goldberg, Sarah R.
; APPLICANT: Caparon, Mairé H.
; APPLICANT: Easton, Alan M.
; APPLICANT: Klein, Barbara K.
; APPLICANT: McKearn, John P.
; APPLICANT: Orlins, Peter O.
; APPLICANT: Paik, Kumhan
; APPLICANT: Thomas, John W.
; TITLE OF INVENTION: Multivariant IL-3 Hematopoiesis
; TITLE OF INVENTION: Fusion Protein
; NUMBER OF SEQUENCES: 197
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dennis A. Bennett, G.D. Searle & Co.,

```

ADDRESSEE: Corporate Patent Dept.
STREET: P. O. Box 5110
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60680
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,872A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/192,325
FILING DATE: 14-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Dennis A.
REGISTRATION NUMBER: 34,547
REFERENCE/DOCKET NUMBER: C-2790/1
TELEPHONE: (314)737-6986
TELEFAX: (314)737-6972
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 966 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-446-872A-70

Query Match 19.38; Score 26.8; DB 4; Length 966;
Best Local Similarity 50.08; Pred. No. 5.1;
Matches 67; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
Qy 3 GCCCGGGCTCCACGCTTCTGCTGAGTGAACATACATACAGAGCGCGGAACG 62
Db 522 GCCCTGGAICTCTCACTTCTTAAGCACTTGAGCAGGAGCTCTGGGCGAGGAGCT 463
Qy 63 GGGCGGGAGGAGGAGCAGCAGGCTTTGACCGATAGTAACTCTGCGCTCGGTGCAGC 122
Db 462 GGCAGGCGCTAATGCTGAGCCATCTTTGGAGATTATGAGATTCTTTAGACGGAGGAGA 403
Qy 123 CGAATCTATAAAG 136
Db 402 CGGCTGATAGTAG 389

RESULT 15
US-08-446-872A-71/C
Sequence 71, Application US/08446872A
Patent No. 6361977
GENERAL INFORMATION:
APPLICANT: Abrams, Mark A.
APPLICANT: Bauer, S. C.
APPLICANT: Bradford-Goldberg, Sarah R.
APPLICANT: Caparon, Mair H.
APPLICANT: Easton, Alan M.
APPLICANT: Klein, Barbara K.
APPLICANT: McKearn, John P.
APPLICANT: Ollins, Peter O.
APPLICANT: Paik, Kuman
APPLICANT: Thomas, John W.
TITLE OF INVENTION: Multivariant IL-3 Hematopoiesis
TITLE OF INVENTION: Fusion Protein
NUMBER OF SEQUENCES: 197
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dennis A. Bennett, S.D. Searle & Co.,
ADDRESSEE: Corporate Patent Dept.
STREET: P. O. Box 5110

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60680
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,872A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/192,325
FILING DATE: 14-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Bennett, Dennis A.
REGISTRATION NUMBER: 34,547
REFERENCE/DOCKET NUMBER: C-2790/1
TELEPHONE: (314)737-6986
TELEFAX: (314)737-6972
INFORMATION FOR SEQ ID NO: 71:
SEQUENCE CHARACTERISTICS:
LENGTH: 966 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-446-872A-71

Query Match 19.38; Score 26.8; DB 4; Length 966;
Best Local Similarity 50.08; Pred. No. 5.1;
Matches 67; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
Qy 3 GCCCGGGCTCCACGCTTCTGCTGAGTGAACATACATACAGAGCGCGGAACG 62
Db 522 GCCCTGGAICTCTCACTTCTTAAGCACTTGAGCAGGAGCTCTGGGCGAGGAGCT 463
Qy 63 GGGCGGGAGGAGGAGCAGCAGGCTTTGACCGATAGTAACTCTGCGCTCGGTGCAGC 122
Db 462 GGCAGGCGCTAATGCTGAGCCATCTTTGGAGATTATGAGATTCTTTAGACGGAGGAGA 403
Qy 123 CGAATCTATAAAG 136
Db 402 CGGCTGATAGTAG 389

Search completed: February 3, 2003, 21:03:28
Job time : 53.524 secs


```
Query Match          20.6%; Score 28.6; DB 9; Length 975;
Best Local Similarity 52.0%; Pred. No. 2.1;
Matches 64; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 4 CCGCGCTCACGCTTCTGCTGAGTGAACATACATAACAGAGCGCGGACGG 63
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 682 CTCACACCTCTCTCTGCTCGAGGAGTAAACACTCAATCGGTGGCTGTAGGAG 623

QY 64 GCGGGGAGGAGGAGGAGGAGGAGGCTTACCATAGTAACTCTCGGTGCGTGAGGC 123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 622 GAGCGGAGGAGGAGGAGGAGGATCCTGAGAGAGATAGATCTCGAACCTTCCCTGAGAG 563

QY 124 GAA 126
    ||
Db 562 TAA 560

RESULT 6
US-09-294-093B-6019
; Sequence 6019, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalquidi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 6019
; LENGTH: 280
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 70032861H1
; LOCATION: 279
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-6019

Query Match          19.9%; Score 27.6; DB 10; Length 280;
Best Local Similarity 63.8%; Pred. No. 3.1;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 21 TTCTGCTGAGTGACTGAACATACATAACAGAGCGCGGGAACGGGGGAGGAGGAGA 80
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 139 TCCGCTGAGAGATCTGCACCATAGTGGACAGCGACGACGAGAGAGGCTGTGAGGGTGGCGA 198

QY 81 GCACAG 86
    |||
Db 199 TGACGG 204

RESULT 7
US-09-796-692-6574
; Sequence 6574, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
```

```
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6574
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (341)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (362)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (385)
; OTHER INFORMATION: n=A,T,C or G
US-09-796-692-6574

Query Match          19.7%; Score 27.4; DB 9; Length 393;
Best Local Similarity 53.2%; Pred. No. 4;
Matches 58; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 2 GCGCCCGCTCCAGCTGCTTCTGCTGAGTGAACACTACATAACAGAGCGCGGGAAC 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 GCGCCCGGATGCCCGCAGCTGTCTCAGAGTGGAGTGCACAGACCTGAGGAAGAGCT 240

QY 62 GGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 110
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 241 GGAGAGAGTGGCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 289

RESULT 8
US-10-044-090-30/c
; Sequence 30, Application US/10044090
; Patent No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 30
; LENGTH: 1466
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```



```

; LENGTH: 1622 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 51..1283
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-900-220C-7

Query Match
Best Local Similarity 19.4%; Score 27; DB 8; Length 1622;
Matches 63; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 16 GTGCTTCTGCTGAGTCAGTCGAATCATATAACAGAGCGCGGGAACGGGCGGGGAGGAG 75
Db 1226 GCGTCTCTGCTAGAGAGGCGACGCTCCACCCACTGGGATGTCGGGGCGAGGAGCTG 1285
QY 76 GCAGAGCAGAGCGTTTGACCGATAGTAACTCTGCGCTCGGTGCGAGCGCGAATCTATAAA 135
Db 1286 AAGGACTCCACCGCTGCGCTCTGGAAGTCTGTAAGTGGTCCAGAGCGCTCTCAGCCA 1345
QY 136 GGA 138
Db 1346 GGA 1348

RESULT 12
US-09-151-999-7
; Sequence 7, Application US/09151999
; Patent No. US20020151460A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Elizabeth
; TITLE OF INVENTION: REGULATION OF EPITHELIAL TISSUE BY HEDGEHOG-LIKE
; FILE REFERENCE: CNV-031-02
; CURRENT APPLICATION NUMBER: US/09/151,999
; EARLIER FILING DATE: 1998-08-11
; EARLIER APPLICATION NUMBER: 08/955,552
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: Patentin ver. 2.0
; SEQ ID NO 7
; LENGTH: 1622
; TYPE: DNA
; ORGANISM: Homo sapien Ihh
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (51)..(1283)
US-09-151-999-7

Query Match
Best Local Similarity 19.4%; Score 27; DB 10; Length 1622;
Matches 63; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 16 GTGCTTCTGCTGAGTCAGTCGAATCATATAACAGAGCGCGGGAACGGGCGGGGAGGAG 75
Db 1226 GCGTCTCTGCTAGAGAGGCGACGCTCCACCCACTGGGATGTCGGGGCGAGGAGCTG 1285
QY 76 GCAGAGCAGAGCGTTTGACCGATAGTAACTCTGCGCTCGGTGCGAGCGCGAATCTATAAA 135
Db 1286 AAGGACTCCACCGCTGCGCTCTGGAAGTCTGTAAGTGGTCCAGAGCGCTCTCAGCCA 1345
QY 136 GGA 138
Db 1346 GGA 1348

RESULT 13
US-09-874-503-17
; Sequence 17, Application US/09874503
; Patent No. US20020177188A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Jian
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Li, Hanzhong
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Hymowitz, Sarah G.
; APPLICANT: Tumas, Daniel
; APPLICANT: Starovasnik, Melissa A.
; APPLICANT: VanLoeken, Menno
; APPLICANT: Vandien, Richard
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Yansura, Daniel
; TITLE OF INVENTION: IL-17 HOMOLOGOUS POLYPEPTIDES AND THERAPEUTIC USES THEREOF
; FILE REFERENCE: PL381R1C1P3(US)
; CURRENT APPLICATION NUMBER: US/09/874,503
; CURRENT FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/253,646
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US 60/244,072
; PRIOR FILING DATE: 2000-10-26
; PRIOR APPLICATION NUMBER: US 60/242,837
; PRIOR FILING DATE: 2000-10-24
; PRIOR APPLICATION NUMBER: US 60/175,481
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: US 60/191,007
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/213,807
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: US 60/172,096
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: US 60/138,387
; PRIOR FILING DATE: 1999-06-09
; PRIOR APPLICATION NUMBER: US 60/134,257
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: US 60/131,022
; PRIOR FILING DATE: 1999-04-26
; PRIOR APPLICATION NUMBER: US 60/130,232
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: US 60/113,621
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/085,579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: US 09/854,208
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 09/854,280
; PRIOR FILING DATE: 2001-05-20
; PRIOR APPLICATION NUMBER: US 09/616,744
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: US 09/747,259
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 09/644,848
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: US 09/380,142
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: US 09/380,138
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: US 09/311,832
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: US PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US PCT/US00/34956
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: US PCT/US00/30873
; PRIOR FILING DATE: 2000-11-10
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 16:38:05 ; Search time 715.76 seconds
(without alignments)
10165.006 Million cell updates/sec

Title: US-09-596-141c-3_COPY_1394_1643

Perfect score: 250

Sequence: 1 gggccggcggtccacgtgct.....caggcgagggcgaggagtc 250

Scoring table: IDENTITY NUC

Gapop 10.C , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmul:*

1: gb.ba:*

2: gb.htg:*

3: gb.in:*

4: gb.om:*

5: gb.ov:*

6: gb.pat:*

7: gb.pl:*

8: gb.pr:*

9: gb.ro:*

10: gb.sts:*

11: gb.un:*

12: gb.vi:*

13: gb.vi:*

14: gb.vi:*

15: em.ba:*

16: em.fun:*

17: em.fun:*

18: em.in:*

19: em.in:*

20: em.in:*

21: em.in:*

22: em.ov:*

23: em.pat:*

24: em.ph:*

25: em.pl:*

26: em.ro:*

27: em.sts:*

28: em.un:*

29: em.vi:*

30: em.htg_hum:*

31: em.htg_in:*

32: em.htg_other:*

33: em.htg_mus:*

34: em.htg_pln:*

35: em.htg_rod:*

36: em.htg_mam:*

37: em.htg_vrt:*

38: em_sy:*

39: em.htg_hum:*

40: em.htg_mus:*

41: em.htg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	250	100.0	1167	9	HS252201	Homo sapi
2	250	100.0	1643	6	AX060715	Sequence
3	250	100.0	1643	6	AX060894	Sequence
4	250	100.0	69570	2	AC021246	Homo sapi
5	250	100.0	96717	9	AL359182	Human DNA
6	250	100.0	175064	2	AC012230	Homo sapi
7	250	100.0	183999	6	AX092589	Sequence
8	248.4	99.4	3231	6	AX351029	Sequence
9	248.4	99.4	149034	9	AF275948	Homo sapi
10	246.8	98.7	1167	9	AF258623S1	Homo sapi
11	246.8	98.7	201144	9	AF287262	Homo sapi
12	226	90.4	69570	2	AC021246	Homo sapi
13	207.4	83.0	90598	2	AC021345	Homo sapi
14	173	69.2	697	9	AF258627	Homo sapi
15	157.4	63.0	2893	6	AX351031	Sequence
16	132.8	53.1	186889	2	AL807243	Mus muscu
17	132.8	53.1	278572	10	AF287263	Rattus no
18	130.6	52.2	145833	2	AC125837	Sequence
19	98	39.2	7260	6	AX253452	Sequence
20	91	36.4	221	6	AX351032	Sequence
21	89.4	35.8	1556	9	AK024328	Homo sapi
22	89	35.5	1750	9	AK022254	Homo sapi
23	75	30.0	9854	6	AX127831	Sequence
24	75	30.0	9854	6	AX139818	Sequence
25	67	26.8	10442	6	AX060713	Sequence
26	67	26.8	10442	6	AX060892	Sequence
27	67	26.8	10442	9	AF285167	Homo sapi
28	60	24.0	10474	6	AX060719	Sequence
29	60	24.0	10474	6	AX060721	Sequence
30	60	24.0	10474	6	AX060898	Sequence
31	60	24.0	10474	6	AX060900	Sequence
32	39	15.6	207959	2	AC129157	Rattus no
33	37.4	15.0	1599	10	RATMCSR	L27081 Rat melanoc
34	37.4	15.0	75499	9	HSJ955G21	AL121772 Human DNA
35	37.4	15.0	120166	9	AC091390	Homo sapi
36	37.4	15.0	125108	9	AF030453	Homo sapi
37	37.4	15.0	150542	2	AC097603	Rattus no
38	36.4	14.6	38	6	AX092823	Sequence
39	36.4	14.6	129025	2	AC105817	Rattus no
40	36.2	14.5	700	6	AX182501	Sequence
41	36.2	14.5	38796	9	AC118464	Homo sapi
42	36.2	14.5	122557	2	AC011429	Homo sapi
43	36.2	14.5	191397	2	AC023861	Homo sapi
44	36.2	14.5	201635	10	AL611931	Mouse DNA
45	36.2	14.5	220087	2	AL662818	Mus muscu

ALIGNMENTS

RESULT 1
HS252201
LOCUS
DEFINITION
Homo sapiens partial ABC-1 gene for ATP-binding cassette transporter-1, 5'UTR and promoter region.
ACCESSION
AJ252201
VERSION
AT252201.1
KEYWORDS
ABC-1 gene; ATP-binding cassette transporter-1; promoter.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1167)
Porsch-Ozcurumez, M., Langmann, T., Heimerl, S., Borsukova, H.,
Kaminski, W.E., Drobnik, W., Hoyer, C., Schumacher, C. and Schmitz, G.
1167 bp DNA linear PRI 10-APR-2001

TITLE The zinc finger protein 202 (ZNF202) is a transcriptional repressor of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene expression and a modulator of cellular lipid efflux

J. Biol. Chem. 276 (15), 12427-12433 (2001)

21192304 PUBLISHED

11279031

2 (bases 1 to 1167)

Porsch-Oezcuernomez, M.K.

Direct Submission

Submitted (05-JAN-2000) Porsch-Oezcuernomez M.K., Institute for

Clinical Chemistry, University of Regensburg,

FRANZ-Josef-Strauss-Allee 11, 93042 Regensburg, GERMANY

Location/Qualifiers

1..1167

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/map="9q22-31"

/cell_type="leukocyte"

1..1167

/gene="ABC-1"

1..942

/gene="ABC-1"

/function="cholesterol efflux regulatory protein"

896..900

/gene="ABC-1"

1148..1167

/gene="ABC-1"

/function="cholesterol efflux regulatory protein"

BASE COUNT 278 a 315 c 327 g 247 t

ORIGIN

Query Match 100.0%; Score 250; DB 9; Length 1167;

Best Local Similarity 100.0%; Pred. No. 7e-58;

Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCCGGCTCCACGTCGCTTTCTGCTGAGTGAACACATATAACAGAGCCGGGAA 60

|||||

Db 768 GGGCCCCGGCTCCACGTCGCTTTCTGCTGAGTGAACACATATAACAGAGCCGGGAA 827

QY 61 GGGGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120

|||||

Db 828 GGGGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 887

QY 121 GCCGAATCTATAAAGGAAGTACGCCGCAAAACCCCGTAATTCGGAGGAGAGTGAG 180

|||||

Db 888 GCCGAATCTATAAAGGAAGTACGCCGCAAAACCCCGTAATTCGGAGGAGAGTGAG 947

QY 181 TGGGGCGGAGACCCGAGAGCCGAGCCGAGCCCTTCTCTCCGGGCTGCGGAGGCGAGGG 240

|||||

Db 948 TGGGGCGGAGACCCGAGAGCCGAGCCGAGCCCTTCTCTCCGGGCTGCGGAGGCGAGGG 1007

QY 241 CGGGGAGCTC 250

|||||

Db 1008 CGGGGAGCTC 1017

RESULT 2

AX060715

LOCUS AX060715 1643 bp DNA linear PAT 22-JAN-2001

DEFINITION Sequence 3 from Patent WO0078972.

ACCESSION AX060715

VERSION AX060715.1 GI:12406104

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1643)

LAWN, R.M., WIDE, D. and GARVIN, M.

Regulation with binding cassette transporter protein abcl

PATENT: WO 0078972-A 3 28-DEC-2000;

JOURNAL CV THERAPEUTICS, INC. (US)

FEATURES

source Location/Qualifiers
1..1643
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 370 a 413 c 457 g 403 t

ORIGIN

Query Match 100.0%; Score 250; DB 6; Length 1643;

Best Local Similarity 100.0%; Pred. No. 6.8e-58;

Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCCGGCTCCACGTCGCTTTCTGCTGAGTGAACATATAACAGAGCCGGGAA 60

|||||

Db 1394 GGGCCCCGGCTCCACGTCGCTTTCTGCTGAGTGAACATATAACAGAGCCGGGAA 1453

QY 61 GGGGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120

|||||

Db 1454 GGGGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1513

QY 121 GCCGAATCTATAAAGGAAGTACGCCGCAAAACCCCGTAATTCGGAGGAGAGTGAG 180

|||||

Db 1514 GCCGAATCTATAAAGGAAGTACGCCGCAAAACCCCGTAATTCGGAGGAGAGTGAG 1573

QY 181 TGGGGCGGAGACCCGAGAGCCGAGCCGAGCCCTTCTCTCCGGGCTGCGGAGGCGAGGG 240

|||||

Db 1574 TGGGGCGGAGACCCGAGAGCCGAGCCGAGCCCTTCTCTCCGGGCTGCGGAGGCGAGGG 1633

QY 241 CGGGGAGCTC 250

|||||

Db 1634 CGGGGAGCTC 1643

RESULT 3

AX060894

LOCUS AX060894 1643 bp DNA linear PAT 22-JAN-2001

DEFINITION Sequence 3 from Patent WO0078971.

ACCESSION AX060894

VERSION AX060894.1 GI:12406271

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1643)

LAWN, R.M., WIDE, D., ORAM, J.F. and GARVIN, M.

ATP binding cassette transporter protein abcl polypeptides

PATENT: WO 0078971-A 3 28-DEC-2000;

JOURNAL CV THERAPEUTICS, INC. (US)

FEATURES

source Location/Qualifiers

1..1643

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 370 a 413 c 457 g 403 t

ORIGIN

Query Match 100.0%; Score 250; DB 6; Length 1643;

Best Local Similarity 100.0%; Pred. No. 6.8e-58;

Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGCCCCGGCTCCACGTCGCTTTCTGCTGAGTGAACATATAACAGAGCCGGGAA 60

|||||

Db 1394 GGGCCCCGGCTCCACGTCGCTTTCTGCTGAGTGAACATATAACAGAGCCGGGAA 1453

QY 61 GGGGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120

|||||

Db 1454 GGGGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1513

QY 121 GCCGAATCTATAAAGGAAGTACGCCGCAAAACCCCGTAATTCGGAGGAGAGTGAG 180

|||||

Db 1514 GCCGAATCTATAAAGGAAGTACGCCGCAAAACCCCGTAATTCGGAGGAGAGTGAG 1573

QY 181 TGGGGCGGAGACCCGAGAGCCGAGCCGAGCCCTTCTCTCCGGGCTGCGGAGGCGAGGG 240

|||||

```

Db 1574 TGGGGCGGACCGACGAGCGGACCGCTCTCTCCGGGCTCGGCGAGCGCAGG 1633
Qy 241 CGGGGAGGTC 250
Db 1634 CGGGGAGGTC 1643
|||||
Db 1634 CGGGGAGGTC 1643

RESULT 4
AC021246
LOCUS Homo sapiens clone RP11-1N10, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC021246
ACCESSION AC021246
VERSION AC021246.2 GI:9119882
KEYWORDS HTG; HTGS_PHASES;
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 59570)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Hirotsune, S., Baldwin, J., Barna, N., Becker, R., Bida, F.,
Boguslavsky, L., Bouckheiser, B., Brown, A., Burkett, G., Castle, A.,
Chapel, X., Cielango, M., Collins, S., Collymore, A., Cooke, P.,
DeRellano, K., Dewar, K., Domako, K., Doyle, M., Feneclor, J.,
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardina, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehotzky, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McCut, A., McKernan, K.,
McNeeley, R., Meltrich, J., Meneus, L., Morrow, J., Nay, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testave, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6705871.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- genome center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2512
Center clone name: L_N10
-----
* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 871: contig of 871 bp in length
872 971: gap of 100 bp
972 1834: contig of 863 bp in length
1835 1934: gap of 100 bp
1935 2804: contig of 870 bp in length
2805 2904: gap of 100 bp
2905 3745: contig of 841 bp in length
3746 3845: gap of 100 bp
3846 4696: contig of 851 bp in length
4697 4796: gap of 100 bp
4797 5640: contig of 844 bp in length
5641 5740: gap of 100 bp
5741 5540: contig of 800 bp in length
5541 6640: gap of 100 bp
6641 7509: contig of 869 bp in length
7510 7609: gap of 100 bp
7610 8479: contig of 870 bp in length
8480 8579: gap of 100 bp
8580 9430: contig of 851 bp in length
9431 9530: gap of 100 bp
9531 10376: contig of 846 bp in length
10377 10476: gap of 100 bp
10477 11322: contig of 846 bp in length
11323 11422: gap of 100 bp
11423 12302: contig of 880 bp in length
12303 12402: gap of 100 bp
12403 13280: contig of 878 bp in length
13281 13380: gap of 100 bp
13381 14241: contig of 861 bp in length
14242 14341: gap of 100 bp
14342 15196: contig of 855 bp in length
15197 15296: gap of 100 bp
15297 16123: contig of 827 bp in length
16124 16223: gap of 100 bp
16223 17072: contig of 849 bp in length
17073 17172: gap of 100 bp
17173 18041: contig of 869 bp in length
18042 18141: gap of 100 bp
18142 19009: contig of 858 bp in length
19010 19109: gap of 100 bp
19110 19966: contig of 857 bp in length
19967 20066: gap of 100 bp
20067 20921: contig of 855 bp in length
20922 21021: gap of 100 bp
21022 21865: contig of 844 bp in length
21865 21965: gap of 100 bp
21966 22832: contig of 867 bp in length
22833 22932: gap of 100 bp
22933 23780: contig of 848 bp in length
23781 23880: gap of 100 bp
23881 24733: contig of 853 bp in length
24734 24833: gap of 100 bp
24834 25670: contig of 837 bp in length
25671 25770: gap of 100 bp
25771 26621: contig of 851 bp in length
26622 26721: gap of 100 bp
26722 27576: contig of 855 bp in length
27577 27676: gap of 100 bp
27677 28532: contig of 856 bp in length
28533 28632: gap of 100 bp
28633 29492: contig of 860 bp in length
29493 29592: gap of 100 bp
29593 30455: contig of 863 bp in length
30456 30555: gap of 100 bp
30556 31410: contig of 855 bp in length
31411 31510: gap of 100 bp
31511 32368: contig of 858 bp in length
32369 32468: gap of 100 bp
32469 33312: contig of 844 bp in length
33313 33412: gap of 100 bp
33413 34268: contig of 856 bp in length
34269 34368: gap of 100 bp
34369 35204: contig of 836 bp in length
35205 35304: gap of 100 bp
35305 36156: contig of 852 bp in length
36157 36256: gap of 100 bp
36257 37128: contig of 872 bp in length
37129 37228: gap of 100 bp
37229 38083: contig of 855 bp in length

```



```

* 45449 45548: gap of 100 bp
* 45549 48116: contig of 2568 bp in length
* 48117 48216: gap of 100 bp
* 48217 52618: contig of 4402 bp in length
* 52619 52718: gap of 100 bp
* 52719 56592: contig of 3874 bp in length
* 56593 56692: gap of 100 bp
* 56693 59635: contig of 2943 bp in length
* 59636 59735: gap of 100 bp
* 59736 63661: contig of 3926 bp in length
* 63662 63761: gap of 100 bp
* 63762 68437: contig of 4676 bp in length
* 68438 68537: gap of 100 bp
* 68538 71458: contig of 2921 bp in length
* 71459 71558: gap of 100 bp
* 71559 76888: contig of 5330 bp in length
* 76889 76988: gap of 100 bp
* 76989 82113: contig of 5125 bp in length
* 82114 82213: gap of 100 bp
* 82214 88220: contig of 6007 bp in length
* 88221 88320: gap of 100 bp
* 88321 93499: contig of 5179 bp in length
* 93500 93599: gap of 100 bp
* 93600 97501: contig of 4302 bp in length
* 97502 98001: gap of 100 bp
* 98002 103016: contig of 5015 bp in length
* 103017 103116: gap of 100 bp
* 103117 109778: contig of 6062 bp in length
* 109779 109278: gap of 100 bp
* 109279 117307: contig of 8029 bp in length
* 117308 117407: gap of 100 bp
* 117408 124079: contig of 6672 bp in length
* 124080 124179: gap of 100 bp
* 124180 131281: contig of 7102 bp in length
* 131282 131381: gap of 100 bp
* 131382 138059: contig of 6672 bp in length
* 138060 138159: gap of 100 bp
* 138160 145431: contig of 7332 bp in length
* 145492 145591: gap of 100 bp
* 145592 157391: contig of 11800 bp in length
* 157392 157491: gap of 100 bp
* 157492 175064: contig of 17573 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..175064
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="RP11-1M10"
                /clone="RP11-1M10"
            1..13003
                /note="assembly_fragment"
            1104..12634
                /note="assembly_fragment"
            2735..4415
                /note="assembly_fragment"
            4516..5785
                /note="assembly_fragment"
            5866..7879
                /note="assembly_fragment"
            7980..9686
                /note="assembly_fragment"
            9787..12253
                /note="assembly_fragment"
            12354..15228
                /note="assembly_fragment"
            15329..17200
                /note="assembly_fragment"
            17301..20131
                /note="assembly_fragment"
            20232..22587
                /note="assembly_fragment"
            22688..25707
                /note="assembly_fragment"
            25808..26194

```

```

/note="assembly_fragment"
28285..31338
/note="assembly_fragment"
31439..34299
/note="assembly_fragment"
34400..38318
/note="assembly_fragment"
38419..42835
/note="assembly_fragment"
42936..45448
/note="assembly_fragment"
45549..48116
/note="assembly_fragment"
48217..52618
/note="assembly_fragment"
52719..56592
/note="assembly_fragment"
56693..59635
/note="assembly_fragment"
59736..63661
/note="assembly_fragment"
63762..68437
/note="assembly_fragment"
68538..71458
/note="assembly_fragment"
clone_end:SP6
vector_side:left
71559..76888

Query Match      100.0%; Score 250; DB 2; Length 175064;
Best Local Similarity 100.0%; Pred. No. 4.9e-58;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCCCCGGCTCCACGCTCTTCTGCTGAGTGAAGTCACTACATAACAGAGCCGGGAA 60
|||||
Db 3503 GGCCCCGGCTCCACGCTCTTCTGCTGAGTGAAGTCACTACATAACAGAGCCGGGAA 3562
QY 61 CGGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
|||||
Db 3563 CGGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3622
QY 121 GCCGATCTATAAAGAGACTGTCTCCGGCAAAACCCGTAATTCGCGAGAGAGTGAG 180
|||||
Db 3623 GCCGATCTATAAAGAGACTGTCTCCGGCAAAACCCGTAATTCGCGAGAGAGTGAG 3682
QY 181 TGGGGCGGGGAGCCGCGAGAGCCGAGGAGCCGAGCCCTTCTCTCCGGGCTCCGCGAGGAGGAG 240
|||||
Db 3683 TGGGGCGGGGAGCCGCGAGAGCCGAGGAGCCGAGCCCTTCTCTCCGGGCTCCGCGAGGAGGAG 3742
QY 241 CGGGGAGGCTC 250
|||||
Db 3743 CGGGGAGGCTC 3752

RESULT 7
AX092589
LOCUS AX092589 183999 bp DNA linear PAT 21-MAR-2001
DEFINITION Sequence 1 from Patent WO0115676.
ACCESSION AX092589
VERSION AX092589.1 GI:134444547
KEYWORDS human.
SOURCE
    ORGANISM
        Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    REFERENCE
        1 (bases 1 to 183999)
        AUTHORS Hayden,M.R., Brooks-Wilson,A.R., Pimstone,S.N. and Clee,S.M.
        TITLE Compositions and methods for modulating xdl cholesterol and
        JOURNAL triglyceride levels
        PATENT: WO 0115676-A 1 08-MAR-2001;
        FEATURES
            Location/Qualifiers
                source
                    1..183999

```

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 49549 a 37944 c 41170 g 54950 t 386 others
ORIGIN
Query Match 100.0%; Score 250; DB 6; Length 183993;
Best Local Similarity 100.0%; Pred. No. 4.9e-58;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGCCCGGCTCCACGTCCTTCTGCTGAGTGAACATACATAAAGAGGCGCGGAA 60
D5 28631 GGGCCCGGCTCCACGTCCTTCTGCTGAGTGAACATACATAAAGAGGCGCGGAA 28690
QY 61 GGGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
D5 28691 GGGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 28750
QY 121 GCCGAATCTATAAAGAACTAGTCCCGGAAAACCCCGTAAATGGCAGGAGAGTGAG 180
D5 28751 GCCGAATCTATAAAGAACTAGTCCCGGAAAACCCCGTAAATGGCAGGAGAGTGAG 28810
QY 181 TGGGCGCGGAGCCCGCAGAGCGGAGCGGAGCCCTTCTCTCCCGGCTCGGCGAGGCGAGG 240
D5 28811 TGGGCGCGGAGCCCGCAGAGCGGAGCGGAGCCCTTCTCTCCCGGCTCGGCGAGGCGAGG 28870
QY 241 CGGGAGGCTC 250
D5 28871 CGGGAGGCTC 28880
RESULT 8
AX351029 AX351029 3231 bp DNA linear PAT 06-FEB-2002
LOCUS
DEFINITION Sequence 1 from Patent WO0183746.
ACCESSION AX351029
VERSION AX351029.1 GI:18616385
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Rosier-Montus, M.F., Prades, C., Lemoine, C., Naudin, L., Deneffe, P.,
Brewer, B., Duverger, N., Remaley, A. and Santamarina-Fojo, S.
Regulatory nucleic acid sequences of the abcl gene
Patent: WO 0183746-A 1 08-NOV-2001;
Aventis Pharma S.A. (FR)
FEATURES
Location/Qualifiers
1..3231
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 809 a 773 c 876 g 773 t
ORIGIN
Query Match 99.4%; Score 248.4; DB 6; Length 3231;
Best Local Similarity 99.6%; Pred. No. 1.8e-57;
Matches 249; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GGGCCCGGCTCCACGTCCTTCTGCTGAGTGAACATACATAAAGAGGCGCGGAA 60
D5 2735 GGGCCCGGCTCCACGTCCTTCTGCTGAGTGAACATACATAAAGAGGCGCGGAA 2794
QY 61 GGGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
D5 2795 GGGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2854
QY 121 GCCGAATCTATAAAGAACTAGTCCCGGAAAACCCCGTAAATGGCAGGAGAGTGAG 180
D5 2855 GCCGAATCTATAAAGAACTAGTCCCGGAAAACCCCGTAAATGGCAGGAGAGTGAG 2914
QY 181 TGGGCGCGGAGCCCGCAGAGCGGAGCGGAGCCCTTCTCTCCCGGCTCGGCGAGGCGAGG 240
D5 2915 TGGGCGCGGAGCCCGCAGAGCGGAGCGGAGCCCTTCTCTCCCGGCTCGGCGAGGCGAGG 2974
```

```
QY 241 CGGGAGGCTC 250
D5 2975 CGGGAGGCTC 2984
RESULT 9
AF275948 AF275948 149034 bp DNA linear PRI 17-JUL-2000
LOCUS
DEFINITION Homo sapiens ABCA1 (ABCA1) gene, complete cds.
ACCESSION AF275948
VERSION AF275948.1 GI:9247085
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 149034)
Santamarina-Fojo, S., Peterson, K., Knapper, C., Qiu, Y., Freeman, L.,
Cheng, J.F., Osorio, J., Remaley, A., Yang, X.P., Haudenschield, C.,
Prades, C., Chimini, G., Blackmon, E., Francois, T., Duverger, N.,
Rubin, E.M., Rosier, M., Deneffe, P., Fredrickson, D.S. and Brewer, H.B.
Jr.
Complete genomic sequence of the human ABCA1 gene: analysis of the
human and mouse ATP-binding cassette A promoter
Proc. Natl. Acad. Sci. U.S.A. 97 (14), 7987-7992 (2000)
20345099
10884428
2 (bases 1 to 149034)
Santamarina-Fojo, S., Peterson, K.M., Knapper, C.L., Freeman, J.A.,
Remaley, A.T., Yang, X.-P., Haudenschield, C.C., Blackmon, E.E.,
Francois, T.L. and Brewer, H.B. Jr.
Direct Submission
Submitted (08-JUN-2000) Molecular Disease Branch, National
Institutes of Health, National Heart, Lung and Blood Institute,
Bethesda, MD 20892, USA
Location/Qualifiers
1..149034
/organism="Homo sapiens"
/db_xref="taxon:9606"
1454..148034
/gene="ABCA1"
Join(1454..1674,25831..25989,40385..40475,45012..45153,
46423..46541,67729..67850,70831..71007,83960..84052,
89010..89250,91962..92101,92433..92549,96758..96955,
97702..97907,98428..98604,100391..100613,102360..102581,
103642..103846,104951..105064,106862..107033,
108023..108154,109450..109602,109805..109943,
110546..110866,112124..112196,113183..113385,
115053..115101,115287..115410,116806..116954,
118504..118728,119964..120062,123094..123283,
124804..124898,126208..126240,127363..127468,
128943..129017,129539..129708,130939..131116,
133114..133229,133341..133485,134526..134649,
135736..135865,136130..136250,137038..137100,
138008..138114,140469..140610,140982..141116,
142060..142163,142646..142738,143397..143640,
144581..148034)
/gene="ABCA1"
/product="ABCA1"
Join(1454..1674,25831..25923)
/gene="ABCA1"
2406..2678
/rpt_family="Alu"
4240..4509
/rpt_family="Alu"
4598..5270
/rpt_family="Alu"
5563..5839
/rpt_family="Alu"
5615..6877
/rpt_family="Alu"
8800..9046
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
```



```

Db 1355 GGGGGGGGAGGAGGAGGAGACAGGCTTTGACCGATAGTAACCTCTCGGCTCGGTGCA 1414
QY 121 GCCGAATCTATAAAGAACTAGTCCCGGCAAAACCCCGTAATTCGGACGAGAGTGAG 180
Db 1415 GCCGAATCTATAAAGAACTAGTCCCGGCAAAACCCCGTAATTCGGACGAGAGTGAG 1474
QY 181 TGGGCGGGGAGCGGAGAGCGGAGCGGAGCGCTTCTCTCCGGGCTCGGCGAGGCGAGG 240
Db 1475 TGGGCGGGGAGCGGAGAGCGGAGCGGAGCGGAGCGCTTCTCTCCGGGCTCGGCGAGGCGAGG 1534
QY 241 CGGGGAGCTC 250
Db 1535 CGGGGAGCTC 1544

RESULT 10
LOCUS AF258623S1 1167 bp DNA linear PRI 23-JUN-2000
DEFINITION Homo sapiens ATP binding cassette transporter 1 (ABCA1) gene,
promoter and exon 1.
ACCESSION AF258623
VERSION AF258623.2 GI:8677405
KEYWORDS
SEGMENT 1 of 4
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1167)
AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C.,
Acuizerat,B.E., Fielding,C.J. and Kane,J.P.
TITLE Analysis of hABC1 gene 5' end; additional peptide sequence,
promoter region, and four polymorphisms
JOURNAL Biochem. Biophys. Res. Commun. 271 (2000) In press
REFERENCE 2 (bases 224 to 1167)
AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C.,
Acuizerat,B.E., Fielding,C.J. and Kane,J.P.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2000) Cardiovascular Research Institute,
University of California, San Francisco, 505 Parnassus Avenue, San
Francisco, CA 94143-0130, USA
REFERENCE 3 (bases 1 to 1167)
AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C.,
Acuizerat,B.E., Fielding,C.J. and Kane,J.P.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-2000) Cardiovascular Research Institute,
University of California, San Francisco, 505 Parnassus Avenue, San
Francisco, CA 94143-0130, USA
REMARK Sequence update by submitter
COMMENT On Jun 23, 2000 this sequence version replaced gi:7769713.
FEATURES
source
1..1167
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q31"
224..844
/promoter
845..1147
/exon
/organism="ABCA1"
/number="1"

BASE COUNT 278 a 313 c 328 g 248 t
ORIGIN
Query Match 98.7%; Score 246.8; DB 9; Length 1167;
Best Local Similarity 99.2%; Pred. No. 5.2e-57;
Matches 248; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGCGCGGCTCCACGCTGCTTTCGCTGAGTCTGAACCTACATAACACAGCGCGGAA 60
Db 768 GGGCGCGGCTCCACGCTGCTTTCGCTGAGTCTGAACCTACATAACACAGCGCGGAA 827
QY 61 CGGGCGGGGAGGAGGAGGAGGAGGAGGCTTTGACCGATAGTAACCTCTCGGCTCGGTGCA 120

```

```

Db 828 GGGGCGGGGAGGAGGAGGAGGAGGCTTTGACCGATAGTAACCTCTCGGCTCGGTGCA 887
QY 121 GCCGAATCTATAAAGAACTAGTCCCGGCAAAACCCCGTAATTCGGACGAGAGTGAG 180
Db 888 GCCGAATCTATAAAGAACTAGTCTCGGCAAAACCCCGTAATTCGGACGAGAGTGAG 947
QY 181 TGGGCGGGGAGCGGAGAGCGGAGCGGAGCGCTTCTCTCCGGGCTCGGCGAGGCGAGG 240
Db 948 TGGGCGGGGAGCGGAGAGCGGAGCGGAGCGGAGCGCTTCTCTCCGGGCTCGGCGAGGCGAGG 1007
QY 241 CGGGGAGCTC 250
Db 1008 CGGGGAGCTC 1017

RESULT 11
LOCUS AF287262 20144 bp DNA linear PRI 29-APR-2001
DEFINITION Homo sapiens ATP-binding cassette 1 sub-family A member 1 (ABCA1)
and SNAP protein genes, complete cds.
ACCESSION AF287262
VERSION AF287262.1 GI:13876612
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 20144)
AUTHORS Qiu,Y., Cavellier,L., Chiu,S., Yang,X., Rubin,E. and Cheng,J.F.
TITLE Human and mouse ABCA1 comparative sequencing and transgenesis
studies revealing novel regulatory sequences
JOURNAL Genomics 73 (1), 66-76 (2001)
MEDLINE 21251004
PUBMED 11352367
REFERENCE 2 (bases 1 to 20144)
AUTHORS Qiu,Y., Cavellier,L., Chiu,S., Rubin,E. and Cheng,J.-F.
TITLE Direct Submission
JOURNAL Submitted (13-JUL-2000) Genome Science Department, Lawrence
Berkeley National Laboratory, 1 Cyclotron Rd, MS 84-171, Berkeley,
CA 94720, USA
FEATURES
Location/Qualifiers
1..20144
/organism="Homo sapiens"
/db_xref="taxon:9606"
33931..181457
/organism="ABCA1"
join(33931..34151,58319..58476,72893..72986,77520..77661,
78931..79049,101048..101189,104152..104328,117287..117379,
122329..122569,125285..125424,125754..125870,
130079..130276,131023..131228,131749..131925,
133708..133930,135677..135898,136959..137163,
138268..138381,140179..140350,141340..141471,
142777..142919,143123..143260,143963..144183,
145441..145523,146657..146859,148527..148575,
148771..148884,150280..150428,152078..152202,
153438..153536,156568..156757,158278..158372,
159682..159714,160837..160942,162417..162491,
163013..163182,164413..164590,166588..166703,
166815..166959,168505..168628,169715..169844,
170109..170229,171017..171079,171987..172093,
174448..174589,174961..175095,176039..176142,
176635..176717,177376..177619,178560..181457)
/product="ATP-binding cassette 1 sub-family A member 1"
33931..34151
/organism="ABCA1"
/number="1"
58319..58476
/organism="ABCA1"
/number="2"
join(58411..58476,72893..72986,77520..77661,78931..79049,
101048..101189,104152..104328,117287..117379,

```

```

123232..12569..12585..12542..12574..12587C..12587C..
130079..130276..131023..131228..131749..131925..
133708..133930..135677..135959..137163..
138268..138381..140179..140350..141340..141471..
142777..142919..143123..143260..143963..144183..
145441..145513..146657..146859..
148771..148884..150280..150428..
153682..153736..156568..156757..
159368..159714..160837..160942..162417..162491..
163103..163182..164413..164590..166588..166703..
166815..166959..168503..168628..169715..169844..
170109..170229..171017..171079..171987..172093..
174448..174589..174961..175095..176039..176142..
176625..176717..177376..177619..178560..178700)
/gene="ABCA1"
/codon_start=1
/product="Atp-binding cassette 1 sub-family A member 1"
/protein_id="AAK43526.1"
/db_xref="GI:13876613"
/translation="MACRPOIRLLLNKNIIFRRQTCOLLEAVAPLFIILISVRL
SYPPYEQHECFPNKAMPSSAGTLPWVOGIIICNANPCFRPTPGEARCVVGNENKSIY
ARLFSADARLLLYSKQDMLKRVLTQQKKSSNKLQDFLDVNETFSGFLH
NLSLKSTVPMKRLADYILHKVFLQGLHLSLNCSSKSEMIQLGDQVSELCGLP
KEKIAAARVLRNMDLKPRLTINSTSPPSKELABATKILSLQTLAQELFSMR
SWDMROEVFLTNWSSSSTOYQAVSRIVCGHPEGGLKIKSLNWEEDNNYKALF
GGNGTEADATFDNSTPYCND-MKNLESPLSLIWKALKPLLVGKILYTPDPAT
RQYMAENVKTFQELAVFHDLEGMEELSPKLTWTFMNSQENMDLYRMLLDSRDNDWE
QQLDCLDWTQELAVFLAKHPEDVSSNGSVITWREAFNQTQARTISRMPCVNLN
KLEPATVEVILINKSMELLDERKQWAGIVFTGTPGSELEPHVKYKIRNDIDNVERT
NKIKDGVDPGPRADPFDMRYVWVGZAYLQDVVEQALIRVITGTERKTGYVMQMPY
PCYVDDITFLVKSRSMLFMTLAWISVAVIKGVYVEKEARLKTMRIMGLDNLILW
PSWISSILLVYNGVILKGLNLLPYSDPSVVFELSVFAVILQCELTSLFL
SRANLAACGGITFTLLPLVLCVAMQDVGFTLKFASLLSPVAGFGCECFALFE
EQGIGVQNDNLFESPVEDDGRNLATISVSMMLFNFLYGVMTWYLEAVFPQIGLRPW
YFPTKTSYWFGESEDEKSPGNSOKRISIEICMEEPETHLKLVSQNLVKVYRDMKV
AVDGLALNFYEQGITFLCHNGAGKITMTSILTGFPPTSGTAVILGXDISEMSTIR
QNLGVCPOHNVFQWLTVEEHWFVARLKLKSKKHVKAEMQFQMALDGLPSSKLKST
SOLSGMQRKLSVALAFVGGKVVILDSPTAGDVPYSRGIWELLKYROGTITLST
HMDADVLGRIATISHGKLCVCGSSLPJKNOLCTGYLLILVKDVESLSSCRNS
STVYLKREDSVQSSDAGLGDHSDIETLNQSAISNLINKHVSEARLVEDIGHEL
TYVLYEAKKEGAFELPHEIDDRSLDLSIGSYSETTEELFLKVAEEGVDRETS
DCTLPARNRRAFQDKQCLRFEDDADNDSDIDPSEIDLLSGMDGKSYQVK
GWLKOQFQVALLKRLIARRSRKGFQAIVPAVFCIALVFLSVLPFGKYPSELE
LQPMWYQNYFVSDNAPEDTGTLELLNALTKDPGFGTRCMGNPDPOTPOAGEEW
TTAPVPTIMDLFONGTMONPSPACQSSDKIKKMLPVCPPGAGGLP2PCRKONTA
DILQDLTGNSDYLVKTVQILIAKSLANKIWNFEFRYGGPSLQVSNLQALPPQSEVN
DAIKMKKELKAKDSSADRFNLNSGRTMTGLDTNNYKVFNNKGRHAISSFELNVIN
NALIRANIKGENPSYGITAFNHELNLTKQLESEVALMTISDVLVSTCVIFAMSFV
PASVFLTIQERSVKAKHLQFTSGVKPYIYWSNFVMDMCFVVPATVILIFCFOQ
KSYVSTNLPVALLLLYGWSITPLMKVPASVFKIPSTAYVVLTSVNLFIGINGSVA
TEVLELFDNLKNNINDILKSVLIPHPFCLGKGLDMVKNQAMADALERFCENRFSV
PLSWDLVGRNLFAMAVEGVV7ELITVLYQYRFRIPRPVYNAKLSPLNDEDEDVRRERO
RLDGGQNDITELIKELTKIYIRKRRKPAVDRCVGPCECEGLLGVNCAAGSIFPKM
RLGDDITVTRGDALFNKSLNSIHVEQNGYCPQFQDAITELLTGREHVEFFALLRGV
PEKEYKRGWELAKGLVYKGYKAGNYSGGNK3KLSTAMALIGGPPVYFDEPTTG
MDPKARFNLNCAKLVKGRSVLTSHSMECEALCTRMATWNGRFRCLGVSHOHLK
NRFGDGYTVIVRAGSNPDLKPVODFFGLAPFGSYLKEKRNKLOQLPSSLSLARI
FSLQDEKRLHIEDYSVSQ7FLDQVVFVNFANDQSDDDHLKDLHLKNTVVDNAVIT
SFLQDEKRLKESV"
72893..72986
/gene="ABCA1"
/number=3
77520..77661
/gene="ABCA1"
/number=4
78331..79049
/gene="ABCA1"
/number=5
101048..101169
/gene="ABCA1"
/number=6
104152..104328
/gene="ABCA1"
/number=7
117287..117379
/gene="ABCA1"
/number=8
122329..122569
/gene="ABCA1"
/number=9
125285..125424
/gene="ABCA1"
/number=10
125734..125870
/gene="ABCA1"
/number=11
130079..130276
/gene="ABCA1"
/number=12
131023..131228
/gene="ABCA1"
/number=13
131749..131925
/gene="ABCA1"
/number=14
133708..133930
/gene="ABCA1"
/number=15
135677..135959
/gene="ABCA1"
/number=16
137163..137163
/gene="ABCA1"
/number=17
138268..138381
/gene="ABCA1"
/number=18
140179..140350
/gene="ABCA1"
/number=19
141340..141471
/gene="ABCA1"
/number=20
142777..142919
/gene="ABCA1"
/number=21
143123..143260
/gene="ABCA1"
/number=22
143963..144183
/gene="ABCA1"
/number=23
145441..145513
/gene="ABCA1"
/number=24
146657..146859
/gene="ABCA1"
/number=25
148527..148575
/gene="ABCA1"
/number=26
148771..148884
/gene="ABCA1"
/number=27
150280..150428
/gene="ABCA1"
/number=28
152079..152202
/gene="ABCA1"
/number=29
153438..153536
/gene="ABCA1"
/number=30
156568..156757
/gene="ABCA1"

```

```
Query Match          96.7%; Score 246.8; DB 9; Length 201144;
Best Local Similarity 99.2%; Pred. No. 3.7e-57;
Matches 248; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGCCCGCGTCCACGCTCTCTGCTGAGTGACTGAACATACATAAAGAGAGCGCGGAA 60
|||||
Db 33772 GGGCCCGCGGCTCCACGCTCTCTGCTGAGTGACTGAACATACATAAAGAGAGCGCGGAA 33831
|||||

QY 61 CGGCGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
|||||
Db 33832 GGGCGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 33891
|||||

QY 121 GCCGAATCTATTAAGAGTACTAGTCCGGGAAAACCCCGTAATTCGAGGAGAGTGAG 180
|||||
Db 33892 GCCGAATCTATTAAGAGTACTAGTCCGGGAAAACCCCGTAATTCGAGGAGAGTGAG 33951
|||||

QY 181 TGGCGCGGGACCGCAGAGCGGAGCGGACCGCTCTCTCCGGGCTCGGCGAGGCGAGGG 240
|||||
Db 33952 TGGCGCGGGACCGCAGAGCGGAGCGGACCGCTCTCTCCGGGCTCGGCGAGGCGAGGG 34011
|||||

QY 241 CGGCGAGGCTC 250
|||||
Db 34012 CGGCGAGGCTC 34021
|||||

RESULT 12
AC021246/c
LOCUS Homo sapiens clone RP11-1N10, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC021246
ACCESSION AC021246
VERSION AC021246.2 GI:9119882
KEYWORDS HTG; STGS; PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 69570)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 69570)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukigalter,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
Ferrelira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehotzky,J., Levice,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Melgrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Testafaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
JOURNAL Direct Submission
COMMENT Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6705871.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Smit,A.F.A. & Green, P. (1996-1997)
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu/
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2512
Center clone name: L_N_10
```

```
-----
* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be generic-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 871: contig of 871 bp in length
* 872 971: gap of 100 bp
* 972 1834: contig of 863 bp in length
* 1835 1934: gap of 100 bp
* 1935 2804: contig of 870 bp in length
* 2805 2904: gap of 100 bp
* 2905 3745: contig of 841 bp in length
* 3746 3845: gap of 100 bp
* 3846 4696: contig of 851 bp in length
* 4697 4796: gap of 100 bp
* 4797 5640: contig of 844 bp in length
* 5641 5740: gap of 100 bp
* 5741 5540: contig of 800 bp in length
* 5541 6640: gap of 100 bp
* 6641 7509: contig of 869 bp in length
* 7510 7609: gap of 100 bp
* 7610 8479: contig of 870 bp in length
* 8480 8579: gap of 100 bp
* 8580 9430: contig of 831 bp in length
* 9431 9530: gap of 100 bp
* 9531 10376: contig of 846 bp in length
* 10377 10476: gap of 100 bp
* 10477 11322: contig of 846 bp in length
* 11323 11422: gap of 100 bp
* 11423 12302: contig of 880 bp in length
* 12303 12402: gap of 100 bp
* 12403 13280: contig of 878 bp in length
* 13281 13380: gap of 100 bp
* 13381 14241: contig of 861 bp in length
* 14242 14341: gap of 100 bp
* 14342 15196: contig of 855 bp in length
* 15197 15296: gap of 100 bp
* 15297 16123: contig of 827 bp in length
* 16124 16223: gap of 100 bp
* 16224 17072: contig of 849 bp in length
* 17073 17172: gap of 100 bp
* 17173 18041: contig of 869 bp in length
* 18042 18141: gap of 100 bp
* 18142 19009: contig of 868 bp in length
* 19010 19109: gap of 100 bp
* 19110 19966: contig of 857 bp in length
* 19967 20066: gap of 100 bp
* 20067 20921: contig of 855 bp in length
* 20922 21021: gap of 100 bp
* 21022 21865: contig of 844 bp in length
* 21866 21965: gap of 100 bp
* 21966 22832: contig of 867 bp in length
* 22833 22932: gap of 100 bp
* 22933 23786: contig of 848 bp in length
* 23781 23880: gap of 100 bp
* 23881 24733: contig of 853 bp in length
* 24734 24833: gap of 100 bp
* 24834 25670: contig of 837 bp in length
* 25671 25770: gap of 100 bp
* 25771 26621: contig of 851 bp in length
* 26622 26721: gap of 100 bp
* 26722 27576: contig of 855 bp in length
* 27577 27676: gap of 100 bp
* 27677 28532: contig of 856 bp in length
* 28533 28632: gap of 100 bp
* 28633 29492: contig of 860 bp in length
```


http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4483
Center clone name: 24_J_9

* NOTE: This record contains 92 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* over-amp relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1
* 911 1010: gap of 100 bp in length
* 1011 1873: contig of 863 bp in length
* 1874 1973: gap of 100 bp
* 1974 2824: contig of 851 bp in length
* 2825 2924: gap of 100 bp in length
* 2925 3802: contig of 878 bp in length
* 3803 3902: gap of 100 bp
* 3903 4816: contig of 914 bp in length
* 4817 4916: gap of 100 bp
* 4917 5759: contig of 843 bp in length
* 5760 5859: gap of 100 bp in length
* 5860 6764: contig of 905 bp in length
* 6765 6864: gap of 100 bp
* 6865 7747: contig of 883 bp in length
* 7748 7847: gap of 100 bp
* 7848 8755: contig of 908 bp in length
* 8756 8855: gap of 100 bp
* 8856 9753: contig of 898 bp in length
* 9754 9855: gap of 100 bp
* 9854 10757: contig of 904 bp in length
* 10758 10857: gap of 100 bp
* 10858 11732: contig of 875 bp in length
* 11733 11832: gap of 100 bp
* 11833 12739: contig of 907 bp in length
* 12740 12839: gap of 100 bp
* 12840 13710: contig of 871 bp in length
* 13711 13810: gap of 100 bp
* 13811 14684: contig of 874 bp in length
* 14685 14784: gap of 100 bp
* 14785 15662: contig of 878 bp in length
* 15663 15762: gap of 100 bp
* 15763 16677: contig of 915 bp in length
* 16678 16777: gap of 100 bp
* 16778 17678: contig of 901 bp in length
* 17679 17778: gap of 100 bp
* 17779 18679: contig of 901 bp in length
* 18680 18779: gap of 100 bp
* 18780 19632: contig of 853 bp in length
* 19633 19732: gap of 100 bp
* 19733 20634: contig of 902 bp in length
* 20635 20734: gap of 100 bp
* 20735 21620: contig of 886 bp in length
* 21621 21720: gap of 100 bp
* 21721 22579: contig of 859 bp in length
* 22580 22679: gap of 100 bp
* 22680 23568: contig of 889 bp in length
* 23569 23668: gap of 100 bp
* 23669 24554: contig of 886 bp in length
* 24555 24654: gap of 100 bp
* 24655 25521: contig of 867 bp in length
* 25522 25621: gap of 100 bp

* 25622 26487: contig of 866 bp in length
* 26488 26587: gap of 100 bp
* 26588 27464: contig of 877 bp in length
* 27465 27564: gap of 100 bp
* 27565 28466: contig of 902 bp in length
* 28467 28566: gap of 100 bp
* 28567 29464: contig of 898 bp in length
* 29465 29564: gap of 100 bp
* 29565 30447: contig of 883 bp in length
* 30448 30547: gap of 100 bp
* 30548 31453: contig of 906 bp in length
* 31454 31554: gap of 100 bp
* 31554 32452: contig of 899 bp in length
* 32453 32552: gap of 100 bp
* 32553 33447: contig of 895 bp in length
* 33448 33547: gap of 100 bp
* 33548 34435: contig of 888 bp in length
* 34436 34535: gap of 100 bp
* 34536 35433: contig of 898 bp in length
* 35434 35533: gap of 100 bp
* 35534 36440: contig of 907 bp in length
* 36441 36540: gap of 100 bp
* 36541 37422: contig of 882 bp in length
* 37423 37522: gap of 100 bp
* 37523 38402: contig of 880 bp in length
* 38403 38502: gap of 100 bp
* 38503 39380: contig of 878 bp in length
* 39381 39480: gap of 100 bp
* 39481 40407: contig of 927 bp in length
* 40408 40507: gap of 100 bp
* 40508 41405: contig of 898 bp in length
* 41406 41505: gap of 100 bp
* 41506 42399: contig of 894 bp in length
* 42400 42499: gap of 100 bp
* 42500 43357: contig of 858 bp in length
* 43358 43457: gap of 100 bp
* 43458 44356: contig of 899 bp in length
* 44357 44456: gap of 100 bp
* 44457 45325: contig of 869 bp in length
* 45326 45425: gap of 100 bp
* 45426 46305: contig of 860 bp in length
* 46306 46405: gap of 100 bp
* 46406 47302: contig of 897 bp in length
* 47303 47402: gap of 100 bp
* 47403 48293: contig of 891 bp in length
* 48294 48393: gap of 100 bp
* 48394 49280: contig of 887 bp in length
* 49281 49380: gap of 100 bp
* 49381 50250: contig of 870 bp in length
* 50251 50350: gap of 100 bp
* 50351 51209: contig of 859 bp in length
* 51210 51309: gap of 100 bp
* 51310 52221: contig of 912 bp in length
* 52222 52321: gap of 100 bp
* 52322 53205: contig of 884 bp in length
* 53206 53305: gap of 100 bp
* 53306 54225: contig of 920 bp in length
* 54226 54325: gap of 100 bp
* 54326 55206: contig of 881 bp in length
* 55207 55306: gap of 100 bp
* 55307 56185: contig of 859 bp in length
* 56186 56285: gap of 100 bp
* 56286 57124: contig of 859 bp in length
* 57125 57224: gap of 100 bp
* 57225 58121: contig of 897 bp in length
* 58122 58221: gap of 100 bp
* 58222 59075: contig of 854 bp in length
* 59076 59175: gap of 100 bp
* 59176 60058: contig of 883 bp in length
* 60059 60158: gap of 100 bp
* 60159 61067: contig of 909 bp in length
* 61068 61167: gap of 100 bp
* 61168 62051: contig of 884 bp in length

Search completed: February 4, 2003, 01:27:29
Job time : 1021.76 secs

GenCore version 5.1.3
Copyr:ght (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: february 3, 2003, 12:46:48 ; Search time 68.2784 Seconds
(without alignments)
B245.650 Million cell updates/sec

Title: US-09-596-141c-3_COPY_1394_1643

Perfect score: 250

Sequence: 1 cggccggcgtccacgtgct.....caggccaggcgaggagctc 250

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	250	100.0	1197	24	ABL56400 Human large ATP-bi
2	250	100.0	1643	22	AAF24681 Nucleotide sequenc
3	250	100.0	1643	22	AAF24703 Nucleotide sequenc
4	250	100.0	18399	22	AAF92831 Human ABC1 genomic
5	248.4	99.4	3231	24	AAD37265 Human ABC1 transcr
6	174.4	69.8	2910	21	AAC37267 Human ABC1 transcr
7	119.4	47.8	227	21	AAC05615 Human secreted pro
8	98	39.2	7260	22	AAD21326 Human ATP binding
9	98	39.2	7260	22	AAI70315 Human ATP binding

10	91	36.4	221	24	AAD37268 Human ABC1 gene ex
11	89.4	35.8	736	22	AAH07432 Human cDNA clone (
12	89.4	35.8	1556	22	AAH18606 Human cDNA sequenc
13	89	35.6	1750	22	AAH04729 Human cDNA clone (
14	89	35.6	1750	22	AAH17451 Human cDNA sequenc
15	77.2	30.9	7281	22	AAK51683 Human polynucleoti
16	75.4	30.2	7086	22	ABA09200 Human ABCA1 homolo
17	75.4	30.2	7086	22	AAK52667 Human polynucleoti
18	75	30.0	9854	22	AAK06121 Human ABC1 DNA seq
19	67	26.8	10442	22	AAF24680 Nucleotide sequenc
20	67	26.8	10442	22	AAF24702 Nucleotide sequenc
21	60	24.0	10474	22	AAF24685 Nucleotide sequenc
22	60	24.0	10474	22	AAF24686 Nucleotide sequenc
23	60	24.0	10474	22	AAF24707 Nucleotide sequenc
24	60	24.0	10474	22	AAF24708 Rat melanocortin r
25	37.4	15.0	1599	16	AAO97701 ABC1 polymorphism
26	36.4	14.6	38	22	AAF93064 Human inflammatory
27	36.2	14.5	700	22	AAH92242 DNA encoding novel
28	34.4	13.8	669	23	AAK83553 Human cDNA 5'-end
29	34	13.6	524	22	AAK92108 Human cDNA clone r
30	34	13.6	524	22	AAK93547 Human full-length
31	34	13.6	1769	22	AAK94827 Human cDNA clone (
32	33.2	13.3	595	22	AAH04659 Human extracellular
33	33.2	13.3	1235	22	AAH08064 Nucleotide sequenc
34	33.2	13.3	1877	22	AAH77797 Human stomach canc
35	33.2	13.3	1877	22	AAI93858 Human cDNA sequenc
36	33.2	13.3	1877	22	AAH15566 Human aggrecaan deg
37	32.6	13.0	3250	20	AAK00726 Human ORFX ORF90
38	32.6	13.0	5530	21	AAK75435 Aspergillus oryzae
39	31.8	12.7	1048	21	AAFL1080 Human protein p903
40	31.6	12.6	1260	22	AAF28692 Human cancer agent
41	31.4	12.6	1704	22	AAK60816 Melanocortin-5 rec
42	31.2	12.5	975	18	AAK768791 Melanocortin-5 rec
43	31.2	12.5	975	19	AAK62253 Mouse melanocortin
44	31.2	12.5	975	19	AAV03916 Mouse melanocortin
45	31.2	12.5	975	19	AAV06405

ALIGNMENTS

RESULT 1
ABL56400
ID ABL56400 standard; DNA; 1197 BP.

NC ABL56400;

XX 30-JUL-2002 (first entry)

DI Human large ATP-binding cassette transporter 1(hABC1) promoter sequence.

DE Human large ATP-binding cassette transporter 1; ABC1; promoter;

KW Human; large ATP-binding cassette transporter 1; ABC1; promoter;

KW antiarteriosclerotic; gene transfer; transactivator; ds.

XX Homo sapiens.

OS WO200183506-A1.

PN 08-NOV-2001.

PD 27-APR-2001; 2001WO-US13654.

PF 28-APR-2000; 2000US-05560372.

PR (USCO) UNIV COLUMBIA NEW YORK.

XX Tail AR;

XX WPI; 2002-049334/06.

XX Novel isolated human large ATP-binding cassette transporter 1 promoter-
PI capable of directing transcription of heterologous coding sequence
PT positioned downstream to it, useful for expressing foreign DNA in host


```
PT cells -
XX
PS Claim 1; Fig 3; 68pp; English.
XX
CC The invention relates to an isolated human large ATP-binding cassette
CC transporter 1 (ABCI) promoter capable of directing transcription of
CC heterologous coding sequence positioned downstream to it. The ABCI
CC promoter is useful for expressing foreign DNA in a host cell, by
CC introducing into the host cell a gene transfer vector comprising the
CC promoter operably linked to a foreign DNA encoding a desired polypeptide
CC or RNA, where the foreign DNA is expressed. The gene transfer can be
CC introduced into the host cell by adenovirus infection, liposome-mediated
CC transfer, topical application to the cell or microinjection. The gene
CC transfer vector encodes and expresses a reporter molecule. The method
CC further involves introducing into the cell a gene transfer vector
CC comprising a nucleic acid segment encoding a transactivator protein
CC capable of upregulating the ABCI promoter, or contacting the cell with
CC the transactivator protein, or an agonist of the transactivator protein.
CC Modulators of human ABCI gene expression are useful for treating
CC atherosclerosis. The present sequence represents the ABCI promoter.
XX
SQ Sequence 1197 BP; 284 A; 314 C; 328 G; 271 T; 0 other;
Query Match 100.0%; Score 250; DB 24; Length 1197;
Best Local Similarity 100.0%; Pred. No. 1.7e-66;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGCCCCGGCTCCACGTGCTTCTCTGAGTGACTGACTACATAAACAAGAGCGCGGAA 60
DB 932 GGGCCCCGGCTCCACGTGCTTCTCTGAGTGACTGACTACATAAACAAGAGCGCGGAA 991
QY 61 CGGGCGGGGAGGAGACAGCGCTTTCACCGTAGTAGCTCTCGCGCTGGGTGCA 120
DB 992 CGGGCGGGGAGGAGGAGACAGCGCTTTCACCGTAGTAGCTCTCGCGCTGGGTGCA 1051
QY 121 CGCGAATCTATAAAGGAACTAGTCCCGGCAAAAACCCGTAATTCGAGCGAGAGTGAG 180
DB 1052 CGCGAATCTATAAAGGAACTAGTCCCGGCAAAAACCCGTAATTCGAGCGAGAGTGAG 1111
QY 181 TGGGCGCGGGACCCGACGAGCGGACCGACCGCTTCTCTCCGGGCTGGCGGAGGCGAGG 240
DB 1112 TGGGCGCGGGACCCGACGAGCGGACCGACCGCTTCTCTCCGGGCTGGCGGAGGCGAGG 1171
QY 241 CGGGGAGCTC 250
DB 1172 CGGGGAGCTC 1181
RESULT 2
AAF24681
ID AAF24681 standard; DNA; 1643 BP.
XX
AC AAF24681;
XX
DT 20-APR-2001 (first entry)
XX
DE Nucleotide sequence of the 5' flanking region of the human ABCI gene.
XX
KW Human; adenosine triphosphate binding cassette protein 1; ABCI;
KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
KW chromosome 9q22-q31; heart disease; hypercholesterolemia;
KW atherosclerosis; cholesterol transport; ss.
XX
OS Homo sapiens.
XX
PN WO200078972-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WO-US16765.
XX
PR 18-JUN-1999; 99US-0140264.
PR 14-SEP-1999; 99US-0153872.
```

```
PR 19-NOV-1999; 99US-0166573.
XX (CVTH-) CV THERAPEUTICS INC.
XX
PI Lawn RM, Wade D, Garvin M;
XX WPI; 2001-137812/14.
XX
PT Adenosine triphosphate (ATP) binding cassette (ABC) polynucleotide,
PT useful for the development of agents for the treatment of heart disease
PT and other disorders associated with hypercholesterolemia and
PT atherosclerosis -
XX
CC Claim 1; Page 143-144; 2:55pp; English.
XX
CC The present sequence represents the 5' flanking region of the human
CC adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABCI
CC resides in cell membranes and utilises ATP hydrolysis to transport a wide
CC variety of substrates across the plasma membrane. ABCI is a pivotal
CC protein in the apolipoprotein-mediated mobilisation of intracellular
CC cholesterol stores. ABCI is defective in Tangier disease, a genetic
CC disorder characterised by abnormal HDL-cholesterol metabolism. The ABCI
CC gene is localised to chromosome 9q22-q31. The ABCI genes and proteins
CC are useful for developing pharmaceutical agents for the treatment of
CC heart disease and other disorders associated with hypercholesterolemia
CC and atherosclerosis. The genes are useful for developing screening assays
CC to screen for compounds that regulate the expression of genes associated
CC with cholesterol transport. The genes and proteins are also useful for
CC are also useful as diagnostic indicators of cardiovascular disease and
CC other disorders associated with hypercholesterolemia.
XX
SQ Sequence 1643 BP; 370 A; 413 C; 457 G; 403 T; 0 other;
Query Match 100.0%; Score 250; DB 22; Length 1643;
Best Local Similarity 100.0%; Pred. No. 1.8e-66;
Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGGCCCCGGCTCCACGTGCTTCTCTGAGTGACTGACTACATAAACAAGAGCGCGGAA 60
DB 1394 GGGCCCCGGCTCCACGTGCTTCTCTGAGTGACTGACTACATAAACAAGAGCGCGGAA 1453
QY 61 CGGGCGGGGAGGAGGAGACAGCGCTTTCACCGTAGTAGCTCTCGCGCTGGGTGCA 120
DB 1454 CGGGCGGGGAGGAGGAGACAGCGCTTTCACCGTAGTAGCTCTCGCGCTGGGTGCA 1513
QY 121 CGCGAATCTATAAAGGAACTAGTCCCGGCAAAAACCCGTAATTCGAGCGAGAGTGAG 180
DB 1514 CGCGAATCTATAAAGGAACTAGTCCCGGCAAAAACCCGTAATTCGAGCGAGAGTGAG 1573
QY 181 TGGGCGCGGGACCCGACGAGCGGACCGACCGCTTCTCTCCGGGCTGGCGGAGGCGAGG 240
DB 1574 TGGGCGCGGGACCCGACGAGCGGACCGACCGCTTCTCTCCGGGCTGGCGGAGGCGAGG 1633
QY 241 CGGGGAGCTC 250
DB 1634 CGGGGAGCTC 1643
RESULT 3
AAF24703
ID AAF24703 standard; DNA; 1643 BP.
XX
AC AAF24703;
XX
DT 20-APR-2001 (first entry)
XX
DE Nucleotide sequence of the 5' flanking region of the human ABCI gene.
XX
KW Human; adenosine triphosphate binding cassette protein 1; ABCI;
KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
KW chromosome 9q22-q31; heart disease; hypercholesterolemia;
KW atherosclerosis; cholesterol transport; ss.
XX
```

OS Homo sapiens.
 PN WO200078971-A2.
 XX
 PD 28-DEC-2000.
 XX
 PF 16-JUN-2000; 2000WO-US16591.
 XX
 PR 18-JUN-1999; 99US-0140264.
 PR 14-SEP-1999; 99US-0153872.
 PR 19-NOV-1999; 99US-0168573.
 XX
 PA (CVTH-) CV THERAPEUTICS INC.
 PA (UNIW) UNIV WASHINGTON.
 XX
 PI Lawn RM, Wade D, Oram JF, Garvin M;
 XX
 DR WPI; 2001-137811/14.
 XX
 PT Adenosine triphosphate (ATP) binding cassette protein (ABC) 1
 PI polynucleotides and polypeptides, useful for treatment of heart disease
 PI and other disorders associated with hypercholesterolemia and
 PI atherosclerosis
 XX
 PS Disclosure; Page 138-139; 211pp; English.
 XX
 CC The present sequence represents the 5' flanking region of the human
 CC adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABC1
 CC resides in cell membranes and utilizes ATP hydrolysis to transport a wide
 CC variety of substrates across the plasma membrane. ABC1 is a pivotal
 CC protein in the apolipoprotein-mediated mobilisation of intracellular
 CC cholesterol stores. ABC1 is defective in Tangier disease, a genetic
 CC disorder characterised by abnormal HDL-cholesterol metabolism. The ABC1
 CC gene is localised to chromosome 9q22-9q31. The ABC1 genes and proteins
 CC are useful for developing pharmaceutical agents for the treatment of
 CC heart disease and other disorders associated with hypercholesterolemia
 CC and atherosclerosis. The genes are useful for developing screening assays
 CC to screen for compounds that regulate the expression of genes associated
 CC with cholesterol transport. The genes and proteins are also useful for
 CC also useful as diagnostic indicators of cardiovascular disease and
 CC other disorders associated with hypercholesterolemia.
 XX
 SQ Sequence 1643 BP; 370 A; 413 C; 457 G; 403 T; 0 other;
 Query Match 100.0%; Score 250; DB 22; Length 1643;
 Best Local Similarity 100.0%; Pred. No. 1.5e-66;
 Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGGCCCCGGTCCACGTCCTTCTGCTGAGTACACGAGTAAACAGAGCGCGGAA 50
 DB 1394 GGGCCCCGGTCCACGTCCTTCTGCTGAGTACACGAGTAAACAGAGCGCGGAA 1453
 QY 61 GGGGCGGGGAGGAGGAGACACAGGCTTTGACCGATAGTAACCTTCGCGTCGCTGCA 120
 DB 1454 GGGGCGGGGAGGAGGAGACACAGGCTTTGACCGATAGTAACCTTCGCGTCGCTGCA 1513
 QY 121 GCCGAATCTATAAAGGAACCTAGTCCCGGCAAAACCCCGTAATTCGAGGAGAGTGGAG 180
 DB 1514 GCCGAATCTATAAAGGAACCTAGTCCCGGCAAAACCCCGTAATTCGAGGAGAGTGGAG 1573
 QY 181 TGGGCGGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 240
 DB 1574 TGGGCGGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1633
 QY 241 CGGGAGGCTC 250
 DB 1634 CGGGAGGCTC 1643
 RESULT 4
 AAF92831 standard; DNA; 183999 BP.
 XX

AC AAF92831;
 XX
 DT 17-MAY-2001 (first entry)
 XX
 DE Human ABC1 genomic DNA.
 XX
 KW High density lipoprotein-cholesterol; HDL-C; cardiovascular; ABC1; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200115676-A2.
 XX
 PD 08-MAR-2001.
 XX
 PF 01-SEP-2000; 2000WO-IB01492.
 XX
 PR 01-SEP-1999; 99US-0151977.
 PR 15-MAR-2000; 2000US-0526193.
 PR 23-JUN-2000; 2000US-0213958.
 XX
 PA (UYBR-) UNIV BRITISH COLUMBIA.
 PA (XENO-) XENON GENETICS INC.
 XX
 PI Hayden MR, Brooks-Wilson AR, Pimstone SN, Clee SM;
 XX
 DR WPI; 2001-244356/25.
 XX
 PT Treating a lower than normal high density lipoprotein-cholesterol
 PI (HDL-C) level, a higher than normal triglyceride level, or a
 PI cardiovascular disease, by administering a compound that modulates LXR-
 PT or RXR-mediated transcriptional activity
 XX
 PS Claim 8; Fig 1; 317pp; English.
 XX
 CC The present invention relates to a method for treating a patient
 CC diagnosed as having a lower than normal high density
 CC lipoprotein-cholesterol (HDL-C) level, a higher than normal
 CC triglyceride level, or a cardiovascular disease, involving
 CC administering a compound that modulates LXR- or RXR-mediated
 CC transcriptional activity or ABC1 expression or activity.
 CC The LXR gene product may be used in an assay to identify
 CC compounds useful for the treatment of a disease or condition selected a
 CC lower than normal HDL cholesterol level, a higher than normal
 CC triglyceride level, and a cardiovascular disease.
 XX
 SQ Sequence 183999 BP; 49549 A; 37944 C; 41170 G; 54950 T; 386 other;
 Query Match 100.0%; Score 250; DB 22; Length 183999;
 Best Local Similarity 100.0%; Pred. No. 7.6e-66;
 Matches 250; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GGGCCCCGGTCCACGTCCTTCTGCTGAGTACACGAGTAAACAGAGCGCGGAA 60
 DB 28631 GGGCCCCGGTCCACGTCCTTCTGCTGAGTACACGAGTAAACAGAGCGCGGAA 28690
 QY 61 GGGGCGGGGAGGAGGAGACACAGGCTTTGACCGATAGTAACCTTCGCGTCGCTGCA 120
 DB 28651 GGGGCGGGGAGGAGGAGACACAGGCTTTGACCGATAGTAACCTTCGCGTCGCTGCA 28750
 QY 121 GCCGAATCTATAAAGGAACCTAGTCCCGGCAAAACCCCGTAATTCGAGGAGAGTGGAG 180
 DB 28751 GCCGAATCTATAAAGGAACCTAGTCCCGGCAAAACCCCGTAATTCGAGGAGAGTGGAG 28810
 QY 181 TGGGCGGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 240
 DB 28811 TGGGCGGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 28870
 QY 241 CGGGAGGCTC 250
 DB 28871 CGGGAGGCTC 28880
 RESULT 5

AAD37265
 ID AAD37265 standard; DNA: 3231 BP.
 XX
 AC AAD37265;
 XX
 DT 21-AUG-2002 (first entry)
 XX
 DE Human ABC1 transcription regulatory DNA #1.
 XX
 KW Human: ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
 KW cholesterol metabolism; hypercholesterolaemia; antisense therapy; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200183746-A2.
 XX
 PD 08-NOV-2001.
 XX
 PF 02-MAY-2001: 200-WO-EP05488.
 XX
 PR 02-MAY-2000: 2000US-201280P.
 XX
 PA (AVET) AVENTIS PHARMA SA.
 XX
 PI Rosier-Montus M, Prades C, Lemoine C, Naudin L, Deneffe P;
 PI Brewer B, Duverger N, Remaley A, Santamarina-Fojo S;
 XX
 DR WPI: 2002-154404/20.
 XX
 PS Isolated nucleic acid useful for modifying the ATP-binding cassette 1
 PT (ABC1) and screening for candidate modulatory compounds or substances
 PT
 XX
 PS Claim 1: Page 130-131; 152pp; English.
 XX
 CC The invention relates a nucleic acid which is capable of regulating the
 CC transcription of human ATP-binding cassette 1 (ABC1) gene, which is a
 CC casual gene for pathologies linked to a dysfunctioning of cholesterol
 CC metabolism, including diseases such as atherosclerosis. Polynucleotides
 CC of the invention are used to screen candidate molecules or substances
 CC that are capable of modulating the transcription of the ABC1 gene. They
 CC are used in antisense therapy. Compositions comprising sequences of the
 CC invention are used to treat hypercholesterolaemia and atherosclerosis.
 CC The present sequence is human ABC1 transcription regulating DNA.
 XX
 SQ Sequence 3231 3p; 803 A; 773 C; 876 G; 773 T; 0 other;

Query Match 99.4%; Score 248.4; DB 24; Length 3231;
 Best Local Similarity 99.6%; Pred. No. 7e-66;
 Matches 249; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GGGCCCGGGTCCACGTGCTTCTGCTGAGTGACTGAACATACATAACAGAGCGCGGAA 60
 Db |||||
 2735 GGGCCCGGGTCCACGTGCTTCTGCTGAGTGACTGAACATACATAACAGAGCGCGGAA 2794
 QY 61 CGGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
 Db |||||
 2795 GGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2854
 QY 121 GCCCAATCTATAAAGGAAGTACTGCCGCAAAACCCCGTAATTGGAGCGAGAGTGAG 180
 Db |||||
 2855 GCCCAATCTATAAAGGAAGTACTGCCGCAAAACCCCGTAATTGGAGCGAGAGTGAG 2914
 QY 181 TGGGGCGGGAGCCCGGAGAGCCGAGCCCTTCTCTCCCGGGCTCGGCGAGGCGAGGG 240
 Db |||||
 2915 TGGGGCGGGAGCCCGGAGAGCCGAGCCCTTCTCTCCCGGGCTCGGCGAGGCGAGGG 2974
 QY 241 CGGGGAGCTC 250
 Db |||||
 2975 CGGGGAGCTC 2984

RESULT 6

AAD37267
 ID AAD37267 standard; DNA: 2910 BP.
 XX
 AC AAD37267;
 XX
 DT 21-AUG-2002 (first entry)
 XX
 DE Human ABC1 transcription regulatory DNA #3.
 XX
 KW Human: ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
 KW cholesterol metabolism; hypercholesterolaemia; antisense therapy; ds.
 XX
 OS Homo sapiens.
 XX
 PH Key Location/Qualifiers
 FT protein_bind 12..23
 FT /*tag= a
 FT /bound_moiety= "LMO2COM/MYOD"
 FT protein_bind 97..107
 FT /*tag= b
 FT /bound_moiety= "DeltaEFl"
 FT protein_bind 110..125
 FT /*tag= c
 FT /bound_moiety= "S8/NKX2.5"
 FT protein_bind 196..211
 FT /*tag= d
 FT /bound_moiety= "S8"
 FT protein_bind 228..237
 FT /*tag= e
 FT /bound_moiety= "GATA"
 FT protein_bind 399..410
 FT /*tag= f
 FT /bound_moiety= "IK2"
 FT protein_bind 412..420
 FT /*tag= g
 FT /bound_moiety= "LYF1"
 FT protein_bind 528..539
 FT /*tag= h
 FT /bound_moiety= "LMO2COM/MYOD/DeltaEFl"
 FT protein_bind 549..556
 FT /*tag= i
 FT /bound_moiety= "LYF1"
 FT protein_bind 558..568
 FT /*tag= j
 FT /bound_moiety= "DeltaEFl"
 FT protein_bind 590..596
 FT /*tag= k
 FT /bound_moiety= "LNX2.5"
 FT protein_bind 608..620
 FT /*tag= l
 FT /bound_moiety= "NFY/CAAT"
 FT protein_bind 708..715
 FT /*tag= m
 FT /bound_moiety= "MZFl"
 FT protein_bind 723..730
 FT /*tag= n
 FT /bound_moiety= "MZFl"
 FT protein_bind 771..785
 FT /*tag= o
 FT /bound_moiety= "HPH2/SRY/EV11"
 FT protein_bind 803..812
 FT /*tag= p
 FT /bound_moiety= "CREBP1/VP2"
 FT protein_bind 831..837
 FT /*tag= r
 FT /bound_moiety= "NKX2.5"
 FT protein_bind 1076..1089
 FT /*tag= s
 FT /bound_moiety= "GATA"
 FT protein_bind 1173..1188
 FT /*tag= t
 FT /bound_moiety= "LXRalpha/DeltaEFl"
 FT protein_bind 1189..1209

```
FT /tag= u
FT /bound_moiety= "DeltaEFL/DYFL/IX2"
FT 1483..1491
FT /tag= v
FT /bound_moiety= "AP4"
FT 1498..1514
FT /tag= w
FT /bound_moiety= "LMO2-COM/MYOB/deltaEFL/E47"
FT 1524..1545
FT /tag= x
FT /bound_moiety= "ZID/deltaEFL"
FT 1597..1607
FT /tag= y
FT /bound_moiety= "DeltaEFL"
FT 1622..1627
FT /tag= z
FT /bound_moiety= "PPAR"
FT 1632..1637
FT /tag= aa
FT /bound_moiety= "PPAR"
FT 1685..1698
FT /tag= ab
FT /bound_moiety= "USE/NMYC/MYCMAX"
FT 1787..1797
FT /tag= ac
FT /bound_moiety= "DeltaEFL"
FT 1809..1819
FT /tag= ad
FT /bound_moiety= "DeltaEFL"
FT 1822..1833
FT /tag= ae
FT /bound_moiety= "SRY"
FT 1840..1850
FT /tag= af
FT /bound_moiety= "AP1"
FT 1942..1956
FT /tag= ag
FT /bound_moiety= "HNF3beta"
FT 1978..1985
FT /tag= ah
FT /bound_moiety= "NKX2.5"
FT 2008..2016
FT /tag= ai
FT /bound_moiety= "PPAR/NKX2.5/PPAR"
FT 2019..2024
FT /tag= aj
FT /bound_moiety= "PPAR/NKX2.5/PPAR"
FT 2051..2059
FT /tag= ak
FT /bound_moiety= "GATA"
FT 2104..2111
FT /tag= al
FT /bound_moiety= "SOX5"
FT 2114..2152
FT /tag= am
FT /bound_moiety= "SRY/HFH/HNF3beta"
FT 2221..2228
FT /tag= an
FT /bound_moiety= "MZF1"
FT 2234..2249
FT /tag= ao
FT /bound_moiety= "IK2/NFkappaB/CREL"
FT 2259..2272
FT /tag= ap
FT /bound_moiety= "LMO2COM/GATA"
FT 2289..2306
FT /tag= aq
FT /bound_moiety= "MZF1/SRY"
FT 2313..2318
FT /tag= ar
FT /bound_moiety= "PPAR"
FT 2321..2326
FT /tag= as
```

```
FT /bound_moiety= "PPAR"
FT 2335..2342
FT /tag= at
FT /bound_moiety= "MZF1"
FT 2361..2384
FT /tag= au
FT /bound_moiety= "HNF3beta/SRY/EV11"
FT 2426..2433
FT /tag= av
FT /bound_moiety= "MZF1"
FT 2442..2451
FT /tag= aw
FT /bound_moiety= "AP4"
FT 2455..2466
FT /tag= ax
FT /bound_moiety= "SRY"
FT 2491..2498
FT /tag= ay
FT /bound_moiety= "STAT"
FT 2524..2534
FT /tag= az
FT /bound_moiety= "STAT/PPAR"
FT 2536..2541
FT /tag= ba
FT /bound_moiety= "PPAR"
FT 2589..2600
FT /tag= bb
FT /bound_moiety= "AP2"
FT 2610..2617
FT /tag= bc
FT /bound_moiety= "MZF1"
FT 2634..2648
FT /tag= bd
FT /bound_moiety= "LMO2COM/MYOB/E47"
FT 2657..2672
FT /tag= be
FT /bound_moiety= "RREB1"
FT 2680..2696
FT /tag= bf
FT /bound_moiety= "MZF1/CMVB"
FT 2728..2740
FT /tag= bc
FT /bound_moiety= "SP1/GC"
FT 2745..2757
FT /tag= bh
FT /bound_moiety= "USE/NMYC/ARNT"
FT 2758..2773
FT /tag= bi
FT /bound_moiety= "NFE2A1"
FT 2774..2787
FT /tag= bj
FT /bound_moiety= "XFD1/HFH"
FT 2794..2806
FT /tag= bk
FT /bound_moiety= "GC/SP1/MZF1"
FT
```

Query Match 69.8%; Score 174.4; DB 24; Length 2910;

Best Local Similarity 99.4%; Pred. No. 2.7e-43;

Matches 175; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 GGGCCCCGGCTCCACGTCTTCTGCTGAGTGAAGTACATTAACAGAGCGCCGGAA 60
DB 2735 GGGCCCCGGCTCCACGTCTTCTGCTGAGTGAAGTACATTAACAGAGCGCCGGAA 2794
QY 61 GGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
DB 2795 GGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2854
QY 121 GCGGAATCTATAAAGGAAGTACTAGTCCCGCAAAACCCCGTAAATTCGAGCGAGAG 176
DB 2855 GCGGAATCTATAAAGGAAGTACTAGTCCCGCAAAACCCCGTAAATTCGAGCGAGAG 2910
```


XX Human ATP binding cassette transporter 1 (ABCI) cDNA.
 XX
 KW ATP binding cassette transporter 1; ABC1; human; lipid disorder;
 KW cholesterol; cardiovascular disease; inflammatory disease;
 KW antinflammatory; antilipemic; antipsoriatic; dermatological;
 KW Tangier disease; coronary heart disease; diagnosis; gene therapy;
 KW polymorphism; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 321..7106
 FT FT /*tag= a
 FT CDS 501..7106
 FT FT /*tag= b
 FT FT /*note= "alternative open reading frame of AAI70314"
 FT FT replace(976,A)
 FT FT /*tag= c
 FT FT variation replace(1516,C)
 FT FT /*tag= d
 FT FT variation replace(2969,G)
 FT FT /*tag= e
 FT FT variation replace(3836,C)
 FT FT /*tag= f
 XX
 PN EPI136554-A1.
 XX
 XX 26-SEP-2001.
 PD
 XX 24-MAR-2000; 2000EP-0106401.
 XX
 XX 24-MAR-2000; 2000EP-0106401.
 XX
 PA (FARB) BAYER AG.
 XX
 PI Schmitz G, Bodzioch M;
 XX
 DR WPI: 2001-640389/74.
 DR P-PSDB; AAM50228.
 XX
 XX New adenosine triphosphate binding cassette transporter gene
 PT polymorphisms, useful for diagnosing and treating lipid disorders,
 PT cardiovascular diseases and inflammatory diseases
 XX
 PS Disclosure; Page 26-28; 41pp; English.
 XX
 CC The present sequence is that of cDNA encoding the human adenosine
 CC triphosphate (ATP) binding cassette transporter 1 (ABCI) protein
 CC (see AAM50227). The sequence includes an extended open reading
 CC frame (ORF) to that provided by the sequence in AAI70314, using
 CC an alternative ATG codon as initiation codon and thereby adding an
 CC extra 40 N-terminal amino acids to the encoded ABC1 protein (see
 CC AAM50228). The invention provides 4 common polymorphisms in the
 CC ABC1 gene. These were identified by sequencing the ABC1 gene in
 CC different Tangier kindreds. In the variant genes (numbering as in
 CC AAI70314), G is changed to A at position 596, T is changed to C at
 CC position 1136, A is changed to G at position 2569 or G is changed
 CC to C at position 3456, or any combination of these. All of these
 CC polymorphisms alter the amino acid sequence of ABC1 and therefore
 CC may affect its function. The 2 most common polymorphisms (G556A
 CC and A2589G) are both associated with a decreased in vitro ApoA-I
 CC mediated efflux of cholesterol from mononuclear phagocytes, a
 CC feature typical of Tangier disease. 3 Of the variants (G596A,
 CC A2589G and G3456C) are significantly increased in a population of
 CC men having low high density lipoprotein-cholesterol levels and
 CC established coronary heart disease (CHD) relative to CHD-free
 CC control subjects. The use of the provided ABC1 polymorphisms for
 CC the diagnosis and treatment of lipid disorders, cardiovascular
 CC diseases, and inflammatory diseases (e.g. psoriasis, lupus
 CC erythematoses) is claimed. Modulation of ABC1 transcripts or
 CC proteins by antisense or ribozyme technology or RNA decoys is also
 CC claimed.

XX
 SQ Sequence 7260 BP; 1834 A; 1765 C; 1905 G; 1756 T; 0 other;
 Query Match 39.2%; Score 98; DB 22; Length 7260;
 Best Local Similarity 100.0%; Pred. No. 7.4e-20;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 153 AAACCCCGTAATTCGAGCGAGAGTGTGGGCGGAGCCCGAGCGAGCCGACCC 212
 DB 1 AAACCCCGTAATTCGAGCGAGAGTGTGGGCGGAGCCCGAGCGAGCCGACCC 212
 QY 213 TTCTCTCCCGGCTGGCGAGCGAGCGAGCGGCGGAGCTC 250
 DB 61 TTCTCTCCCGGCTGGCGAGCGAGCGGCGGAGCTC 98
 RESULT 10
 AAD37268
 ID AAD37268 standard; DNA; 221 BP.
 XX AAD37268;
 AC AAD37268;
 XX
 DT 21-AUG-2002 (first entry)
 XX
 DE Human ABC1 gene exon 1A.
 XX
 KW Human; ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
 KW cholesterol metabolism; hypercholesterolaemia; antisense therapy;
 KW exon 1A; GS.
 XX
 OS Homo sapiens.
 PN WC200183745-A2.
 XX
 PC 08-NOV-2001.
 XX
 PF 02-MAY-2001; 2001WO-EP05488.
 XX
 PR 02-MAY-2000; 2000US-201280P.
 XX
 XX (AVET) AVENTIS PHARMA SA.
 PA
 PI Rosier-Montus M, Prades C, Lemoine C, Naudin L, Deneffe P;
 PI Brewer B, Duverger N, Remaley A, Santamarina-Fojo S;
 XX
 DR WPI: 2002-154404/20.
 XX
 PT Isolated nucleic acid useful for modifying the ATP-binding cassette 1
 PT (ABCI) and screening for candidate modulatory compounds or substances
 PT
 PS Claim 4; Page 132; 152pp; English.
 XX
 CC The invention relates a nucleic acid which is capable of regulating the
 CC transcription of human ATP-binding cassette 1 (ABCI) gene, which is a
 CC casual gene for pathologies linked to a dysfunctioning of cholesterol
 CC metabolism, including diseases such as atherosclerosis. Polynucleotides
 CC of the invention are used to screen candidate molecules or substances
 CC that are capable of modulating the transcription of the ABC1 gene. They
 CC are used in antisense therapy. Compositions comprising sequences of the
 CC invention are used to treat hypercholesterolaemia and atherosclerosis.
 CC The present sequence is human ABC1 gene exon 1A.
 XX
 SQ Sequence 221 BP; 44 A; 52 C; 73 G; 42 T; 0 other;
 Query Match 35.4%; Score 91; DB 24; Length 221;
 Best Local Similarity 100.0%; Pred. No. 3.6e-18;
 Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 160 GTAATTGCCAGCGAGAGTGTGGGCGGAGCCCGAGCGAGCCGACCCCTTCTC 219
 DB 1 GTAATTGCCAGCGAGAGTGTGGGCGGAGCCCGAGCGAGCCGACCCCTTCTC 219


```

/clone="WAGE:4749735"
/clone_lib="NCL_GAP_Skn4"
/issue_type="squamous cell carcinoma"
/lab_host="DHI0B (T1 phage-resistant)"
/notes="Organ: skin; Vector: pcwv-sport6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Cligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCL_GAP Library."
BASE COUNT      285 a      233 c      244 g      236 t
ORIGIN
Query Match      24.1%; Score 60.2; DB 12; Length 998;
Best Local Similarity 89.0%; Pred. No. 1.1e-06;
Matches 65; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 178 GAGTGGGCGGACCGCAGAGCGCCGACCCCTTCTCCCGGGCTGGCGAGGCA 237
      |||||
DB 1  GTAGTGGGCGGCGACCGCAGAGCGCCGACCCCTTCTCTCCCGGGCTGGCGAGGCA 60

QY 238 GGGCGGGAGCTC 250
      |||||
DB 61 GGGCGGGAGCTC 73

RESULT 7
CNS006N3/3
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR14H04 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL065775
VERSION
AL065775.1 GI:4944655
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Ectopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 912)
AUTHORS
Smith, J., Bruley, C.K., Paton, I.R., Dunn, I., Jones, C.F., Windsor, D.,
Morrice, D.R., Law, A.S., Masabanda, J., Sazanov, A., Waddington, D.,
Fries, R. and Burt, D.W.
Differences in gene density on chicken macrochromosomes and
microchromosomes
Anim. Genet. 31 (2), 96-103 (2000)
MEDLINE
20244064
PUBMED
10782207
REFERENCE
2 (bases 1 to 1863)
AUTHORS
Smith, J., Bruley, C.K., Paton, I.R., Law, A.S., Masabanda, J.,
Waddington, D., Fries, R. and Burt, D.W.
Direct Submission
TITLE
Submitted (12-AUG-1998) Division of Molecular Biology, Roslin
Institute, Roslin, Midlothian EH25 9PS, Scotland, UK
FEATURES
Location/Qualifiers
source
1, 1863
/organism="Callus gallus"
/db_xref="taxon:9031"
/chromosome="microchromosome"
BASE COUNT      408 a      392 c      580 g      482 t
ORIGIN
Query Match      15.1%; Score 37.8; DB 17; Length 1863;
Best Local Similarity 54.7%; Pred. No. 2.8;
Matches 75; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 17 TGTCTTCTGTGACTGACTACATACATACAGAGCGCGGAGCGGGGAGAGAG 76
      |||||
DB 239 TGATTACAGAGTGTACTACTACATAGCAGCGCCACTATTAGCTTTGGCGTAGC 298

QY 77 GAGAGCAGAGCTTTGACCGATAGTAACCTTCGCGTCGTCGACCGGATCTATAAAG 136
      |||||
DB 299 GAAAGAGAAATCTGCAGCTCAITCCAGCCACATCCACCCCTGCGAGGTGAATATAACAG 358

QY 137 GAACTAGTCCCGCAAA 153
      |||||
DB 359 GAATGATCTTAGGTAA 375

RESULT 9
CNS016BR/c
LOCUS
Query Match      15.8%; Score 39.6; DB 17; Length 912;
Best Local Similarity 26.7%; Pred. No. 0.74;
Matches 52; Conservative 67; Mismatches 76; Indels 0; Gaps 0;

```

```

QY 50 GAGCGCGGAACGGGCGGAGGAGGAGGACAGAGCTTTTGACCGATAGTAACCTCTG 109
      |||||
DB 694 GGGGVSVVSGGGSSVVVRVGGGSGGGGVSVMVGMGGGVSVMGMSVAGV 635

QY 110 CGCTGGTGCAGCGGAATCTATAAAGGAAGTACTGTCGGGAAAAACCCCTAATTCGCA 169
      |||||
DB 634 VVGVMGVMGGRGVGMGMVRRRVRVGMVGMSSMCCSCMCGSGVGRGGSGVSGSC 575

QY 170 CGGAGAGTGTAGTGGGCGGCGGACCGGAGCGCGACCTTCTCTCCCGGCTCGC 223
      |||||
DB 574 GMSAGASVMSMGVAGASRAASGMCGCGCCGCCGCCGCCGCCGCCGCCGCCGGA 515

QY 230 GCAGGCGAGGCGGG 244
      |||||
DB 514 GAVASGCASGGGSR 500

RESULT 8
GGA200110
LOCUS
DEFINITION
Gallus gallus anonymous sequence from Cosmid mapping to a
microchromosome (Cosmid 20 - Contig 12), genomic survey sequence.
ACCESSION
AJ231825
VERSION
AJ231825.1 GI:3451643
KEYWORDS
GSS: genome survey sequence.
SOURCE
chicken.
ORGANISM
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 1863)
AUTHORS
Smith, J., Bruley, C.K., Paton, I.R., Dunn, I., Jones, C.F., Windsor, D.,
Morrice, D.R., Law, A.S., Masabanda, J., Sazanov, A., Waddington, D.,
Fries, R. and Burt, D.W.
Differences in gene density on chicken macrochromosomes and
microchromosomes
Anim. Genet. 31 (2), 96-103 (2000)
MEDLINE
20244064
PUBMED
10782207
REFERENCE
2 (bases 1 to 1863)
AUTHORS
Smith, J., Bruley, C.K., Paton, I.R., Law, A.S., Masabanda, J.,
Waddington, D., Fries, R. and Burt, D.W.
Direct Submission
TITLE
Submitted (12-AUG-1998) Division of Molecular Biology, Roslin
Institute, Roslin, Midlothian EH25 9PS, Scotland, UK
FEATURES
Location/Qualifiers
source
1, 1863
/organism="Callus gallus"
/db_xref="taxon:9031"
/chromosome="microchromosome"
BASE COUNT      408 a      392 c      580 g      482 t
ORIGIN
Query Match      15.1%; Score 37.8; DB 17; Length 1863;
Best Local Similarity 54.7%; Pred. No. 2.8;
Matches 75; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 17 TGTCTTCTGTGACTGACTACATACATACAGAGCGCGGAGCGGGGAGAGAG 76
      |||||
DB 239 TGATTACAGAGTGTACTACTACATAGCAGCGCCACTATTAGCTTTGGCGTAGC 298

QY 77 GAGAGCAGAGCTTTGACCGATAGTAACCTTCGCGTCGTCGACCGGATCTATAAAG 136
      |||||
DB 299 GAAAGAGAAATCTGCAGCTCAITCCAGCCACATCCACCCCTGCGAGGTGAATATAACAG 358

QY 137 GAACTAGTCCCGCAAA 153
      |||||
DB 359 GAATGATCTTAGGTAA 375

RESULT 9
CNS016BR
LOCUS
Query Match      15.8%; Score 39.6; DB 17; Length 912;
Best Local Similarity 26.7%; Pred. No. 0.74;
Matches 52; Conservative 67; Mismatches 76; Indels 0; Gaps 0;

```



```

Qy 28 GAGTCAGTAACTACATAAAGACAGAGCGGGAACGGGGCGGGAGGAGGAGACACAGG 87
D5 973 GVGGMGMVGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGM 914
Qy 88 CTTTGACCGATAGTAACCTCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCT 147
D5 913 VVGGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGM 854
Qy 148 GGCAGAAACCCCGTAATTCAGCAGCAGAGTGTAGTGGGCGGCGGAGCCAGCAGCC 207
D5 853 MGCAGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGM 794
Qy 208 GACCTTCCTCCCGCTCGGCGAGGCGAGGCGGCGGCGGCGGCGGCTC 250
D5 793 CACACACCGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGMGM 751

RESULT 14
LOCUS BQ708275/c
DEFINITION AGNCOURT_8475172 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301601
5', mRNA sequence.
ACCESSION BQ708275
VERSION BQ708275.1 GI:21847174
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 967)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. Mark Watson
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M2517 row: 1 column: 18
High quality sequence stop: 539.
FEATURES
Location/Qualifiers
1..967
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6301601"
/lab_host="NIH_MGC_113"
/note="organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 92 a 456 c 259 g 129 t 31 others
ORIGIN
Query Match 13.9%; Score 34.8; DB 14; Length 967;
Best Local Similarity 48.9%; Pred. No. 17;
Matches 87; Conservative 0; Mismatches 91; Indels 3; Gaps 0;

Qy 63 GGGCGGGAGGAGGAGCAGACAGCTTTTACCGATAGTAACCTCTCGCTCGCTCGCT 122
D5 952 GGGCGGGGATAGGAGGCGCGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 893
Qy 123 CGAATCTATAAAGAGACTAGTCCCGGGGAAAAACCCGTAATTCGAGCAGAGTAGTG 182
D5 892 CGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 833

```

```

Qy 183 GGGCGGACCGGACAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
D5 832 GGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 775

RESULT 15
LOCUS BQ673485
DEFINITION AGNCOURT_8189102 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6254855
5', mRNA sequence.
ACCESSION BQ673485
VERSION BQ673485.1 GI:21784319
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1190)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10M2405 row: a column: 10
High quality sequence stop: 226.
FEATURES
Location/Qualifiers
1..1190
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6254865"
/lab_host="NIH_MGC_102"
/tissue_type="epidermoid carcinoma, cell line"
/note="organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 30 a 471 c 581 g 69 t 39 others
ORIGIN
Query Match 13.9%; Score 34.8; DB 14; Length 1190;
Best Local Similarity 48.9%; Pred. No. 18;
Matches 93; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 54 CCGGAAACGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 113
D5 446 CCGGCGCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 505
Qy 114 CGGTGCAGCGCAATCTATATAAGAGCACTAGTCCCGGSCAAAAACCCGTAATTCGAGCG 173
D5 506 GGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 565
Qy 174 GAGTAGTGGGCGGGGAGCCCGCAGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 233
D5 566 GGGGGCGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625
Qy 234 GGCAGGCGCG 243
D5 626 GGCAGGCGCG 635

```


Tue Feb 4 09:39:53 2003

us-09-596-141c-3_copy_1394_1643.rst

Page 8

Search completed: February 4, 2003, 03:27:02
Job time : 487.757 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 12:56:40 : Search time 11:7338 Seconds
(without alignments)
6534.066 Million cell updates/sec

Title: US-09-596-141C-3_COPY_1394_1643

Perfect score: 250

Sequence: 1 gggcccggtccacgtgct.....caggcgagggcgaggagctc 250

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCRU_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	35	14.0	7218	1	US-08-232-463-14
C 2	32.6	13.0	3250	4	US-09-122-126B-14
C 3	31.2	12.5	975	1	US-08-671-525B-9
C 4	31.2	12.5	975	1	US-08-672-109B-9
C 5	31.2	12.5	975	1	US-08-842-045-9
C 6	31.2	12.5	975	2	US-08-842-238-9
C 7	31.2	12.5	975	3	US-08-629-335B-9
C 8	31.2	12.5	978	3	US-08-736-281A-17
C 9	31.2	12.5	978	4	US-09-097-231-17
C 10	30	12.0	966	3	US-08-459-318-68
C 11	30	12.0	966	3	US-08-459-318-70
C 12	30	12.0	966	3	US-08-459-318-71
C 13	30	12.0	966	3	US-08-458-609A-68
C 14	30	12.0	966	3	US-08-458-609A-70
C 15	30	12.0	966	3	US-08-458-609A-71
C 16	30	12.0	966	4	US-08-446-872A-68
C 17	30	12.0	966	4	US-08-446-872A-70
C 18	30	12.0	966	4	US-08-446-872A-71
C 19	30	12.0	966	4	US-08-762-227A-68
C 20	30	12.0	966	4	US-08-762-227A-70
C 21	30	12.0	966	4	US-08-762-227A-71
C 22	30	12.0	966	5	PCT-US95-01185-68
C 23	30	12.0	966	5	PCT-US95-01185-70
C 24	30	12.0	966	5	PCT-US95-01185-71
C 25	30	12.0	1817	2	US-08-870-518-8
C 26	30	12.0	81001	4	US-09-750-580-1
C 27	29.8	11.9	5310	3	US-08-870-126-11

C 28 29.8 11.9 8310 4 US-09-445-247-11
C 29 29.8 11.9 14985 1 US-08-652-972A-6
C 30 29.8 11.9 14985 5 PCT-US96-06231A-6
C 31 29.6 11.8 58750 3 US-09-335-409-1
C 32 29.6 11.8 58750 4 US-09-568-102-1
C 33 29.6 11.8 58750 4 US-05-567-969-1
C 34 29.6 11.8 58750 4 US-09-568-480-1
C 35 29.6 11.8 58750 4 US-09-568-486-1
C 36 29.6 11.8 58750 4 US-09-568-472-1
C 37 29.6 11.8 58750 4 US-09-567-899-1
C 38 29.6 11.8 71989 4 US-09-443-501A-2
C 39 29.4 11.8 1860 3 US-09-165-240-4
C 40 29.4 11.8 1860 4 US-09-568-059-4
C 41 29.4 11.8 4085 3 US-09-165-240-5
C 42 29.4 11.8 4085 4 US-09-568-059-5
C 43 29.4 11.8 11219 1 US-07-642-734C-1
C 44 29.4 11.8 11219 3 US-08-439-009A-1
C 45 28.6 11.4 4403765 4 US-09-103-840A-2

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/232,463
; FILING DATE: 26-AUG-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: US/07/935,313
; FILING DATE: 26-AUG-1991
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,758
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pZipt-Fis
US-08-232-463-14

Query Match 14.0%; Score 35; DB 1; Length 7218;


```
Qy 3 GCCCGGCTCCACGTGCTTCTGCTCAGTGAGCTGAACTACATAAACAGAGAGCGCGGGAACG 62
Db 522 GCCCTGGATCTTCTCACTTGGCTTAAGCACTTGAGCAGGAGGCTCTGGGCGAGGAGCT 463
Qy 63 GGGCGGGGAGGAGGAGAGAGAGAGGCTTTGACCGATAGTAACTCTGCGCTGCGTGCAGC 122
Db 462 GGCAGGGCCCTAATGGGTAGCCCATGTTGGAGATTATGAGATTCTTTAGACCGGAGGAGA 403
Qy 123 CGAATCTATAAAGGAACTAGTCCCGGCAAAACCCCG 160
Db 402 CGGGTTGATAGTAGAGATTGGACCAAGACGTTACCCG 365
```

Search completed: February 3, 2003, 21:02:41
Job time : 25.7338 secs


```
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Steven A.  
; TITLE OF INVENTION: A METHOD FOR ISOLATING AND PURIFYING MULTIPOTENTIAL NEURAL PROGENITOR CELLS  
; FILE REFERENCE: 19603/3580  
; CURRENT APPLICATION NUMBER: US/09/747,810  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/473,003  
; PRIOR FILING DATE: 1999-12-23  
; NUMBER OF SEQ ID NOS: 2  
; SEQ ID NO 1  
; LENGTH: 52216  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-747-810-1
```

	Query Match	Score	DB	Length
Best Local Similarity	51.4%	Pred. No. 5.6;	Mismatches	Conservative
Matches	72;	Indels	Gaps	0;

```
QY 51 AGGCCGGGAACGCGCGAGGAGGAGGCAGCAGAGCTTTGACCAGATAGTAACCTCTGC 110  
Db 30273 AAGTGGCAAGCATTCACGTAGTAGCGGGAGAGCAGTGCCAAAGGCTCTCAGGTAGGACTA 30332  
  
QY 111 GCTCGGTGCACCGCAATCTATAAAGAACAATAGTCCCGGCAAAAACCCCCTAATTGCGAG 170  
Db 30333 GTTGGGAGTATCTCAGGAACCTGAAAGCACGCCACTGTGCGCTGGAGCACTGGGAGGAGAG 30392  
  
QY 171 CGAGAGTCAGTGGGCGCGG _190  
Db 30393 TGAGAGTGGGATGGCCGAGG 30412
```

```
RESULT 6  
US-09-880-107-2168  
; Sequence 2168, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2168  
; LENGTH: 10144  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A L17131  
; NAME/KEY: unsure  
; LOCATION: (1)..(10144)  
; OTHER INFORMATION: n = a or c or g or t  
US-09-880-107-2168
```

	Query Match	Score	DB	Length
Best Local Similarity	48.3%;	Pred. No. 4.3;	Mismatches	Conservative
Matches	86;	Indels	Gaps	0;

```
QY 50 GAGCGCGGAACCGCGCGAGGAGGAGGCACAGGCTTTGACCGNATAGTACCTCTG 109  
Db 8509 GGCGCCAGCCTCTCGGGGTGGAAGAGGGGGCCACCGGCCAGAGCTCACACCAACAATG 8568  
  
QY 110 CGCTCGBTGCAGCGCAATCTATAAAGAACAATAGTCCCGGCAAAAACCCCCTAATTGGGA 169
```

Db	8569	CCCACCTCACAGGAGAAGGAGAGAGGGGCATCTCCGAGGACTCCTCGGAGGAGGAG	8628
Qy	170	GCGAGAGTAGTGGCGCGGACCAGCGAGGCCGAGCCCTTCTCTCCCGGGCTG	227
Db	8629	CAGTGACCCATGCCGCGCCCTGTCTCTCATGTGGAGGACGACTTCCTCTGGGACTG	8686
 RESULT 7 US-09-764-847-1616			
; Sequence 1616, Application US/09764847			
; Patent No. US20020132767A1			
; GENERAL INFORMATION:			
; APPLICANT: Rosen et al.			
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies			
; FILE REFERENCE: PC009			
; CURRENT APPLICATION NUMBER: US/09764,847			
; CURRENT FILING DATE: 2001-01-17			
; Prior application data removed - consult PALM or file wrapper			
; NUMBER OF SEQ ID NOS: 2003			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 1616			
; LENGTH: 10378			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-764-847-1616			
Query Match 12.3% Score 30.8; DB 10; Length 10378;			
Best Local Similarity 47.8%; Pred.No.4.3;			
Matches 89; Conservative 0; Mismatches 97; Indels 0; Gaps 0;			
QY	48	CAGAGGCGCGGAACGGGCGCGGAGGAGGAGCAGACAGAGGCTTTGACCGATAGTAACCTC	107
Db	10095	CTGCGGGTGGCTCGSGACCTGGCCAGCAAGGACAGGAGTGGGCTGGCCAACCTC	10154
QY	108	TGGCTCGGTGCACCGCAATCTATAAAGGAACCTAGTCCCGGCAAACCCCGTAATGC	167
Db	10155	GGGCGCCCTGTGCTGTGCTATGCGGGTCCGAGCGAGGTGGCCGAGCACTACCTCTGGAGGCC	10214
QY	168	GAGGAGAGTAGTCAGTGGGCGCGGACCGCCAGCGAGCGGACCGCTTCTCTCCGGGCTG	227
Db	10215	GTGCGGTGTGTCTCAGAGCTGCCCTCGGGGAGTGTGGCCGGGACTTCACCCAGGTGCTC	10274
QY	228	CGGCAG 233	
Db	10275	CTGCAG 10280	
 RESULT 8 US-09-751-877-1/c			
; Sequence 1, Application US/09751877			
; Patent No. US20020142949A1			
; GENERAL INFORMATION:			
; APPLICANT: Yes, Frances			
; APPLICANT: Denison, Blake			
; APPLICANT: Bour, Barbara			
; APPLICANT: Bihain, Bernard			
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste			
; APPLICANT: Duclert, Aymeric			
; APPLICANT: Boegueleret, Lydie			
; APPLICANT: Ebbeets-Reed, Dana			
; APPLICANT: Salter-Cid, Luisa			
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL			
; FILE REFERENCE: 89.053.REG			
; CURRENT APPLICATION NUMBER: US/09751,877			
; CURRENT FILING DATE: 2000-12-28			
; NUMBER OF SEQ ID NOS: 6			
; SOFTWARE: Patent.pm			
; SEQ ID NO 1			
; LENGTH: 81c01			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			

APPLICANT: SATEL-CIU, LUISA

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH

```

; NAME/KEY: misc_feature
; LOCATION: 10945..12946
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon
; LOCATION: 12947..12958
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 13470..13526
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 13641..13752
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 14271..15968
; OTHER INFORMATION: exon 4
; NAME/KEY: misc_feature
; LOCATION: 15969..17969
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 1239
; OTHER INFORMATION: 20-828-311 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 12347
; OTHER INFORMATION: 17-42-319 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 15241
; OTHER INFORMATION: 17-41-250 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 42219
; OTHER INFORMATION: 20-841-149 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 45442
; OTHER INFORMATION: 20-842-115 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 77058
; OTHER INFORMATION: 20-853-415 : polymorphic base C or T
; NAME/KEY: primer_bind
; LOCATION: 929..949
; OTHER INFORMATION: 20-828.pu
; NAME/KEY: primer_bind
; LOCATION: 1357..1377
; OTHER INFORMATION: 20-828.rp complement
; NAME/KEY: primer_bind
; LOCATION: 12029..12050
; OTHER INFORMATION: 17-42.pu
; NAME/KEY: primer_bind
; LOCATION: 2581..12603
; OTHER INFORMATION: 17-42.rp complement
; NAME/KEY: primer_bind
; LOCATION: 14992..15012
; OTHER INFORMATION: 17-41.pu
; NAME/KEY: primer_bind
; LOCATION: 15460..15482
; OTHER INFORMATION: 17-41.rp complement
; NAME/KEY: primer_bind
; LOCATION: 42070..42090
; OTHER INFORMATION: 20-841.pu
; NAME/KEY: primer_bind
; LOCATION: 42572..42591
; OTHER INFORMATION: 20-841.rp complement
; NAME/KEY: primer_bind
; LOCATION: 45328..45347
; OTHER INFORMATION: 20-842.pu
; NAME/KEY: primer_bind
; LOCATION: 45863..45883
; OTHER INFORMATION: 20-842.rp complement
; NAME/KEY: primer_bind
; LOCATION: 76644..76664
; OTHER INFORMATION: 20-853.pu
; NAME/KEY: primer_bind
; LOCATION: 77166..77185
; OTHER INFORMATION: 20-853.rp complement
; NAME/KEY: primer_bind

```

```

; LOCATION: 1220..1238
; OTHER INFORMATION: 20-828-311.mis
; NAME/KEY: primer_bind
; LOCATION: 1240..1258
; OTHER INFORMATION: 20-828-311.mis complement
; NAME/KEY: primer_bind
; LOCATION: 2328..12346
; OTHER INFORMATION: 17-42-319.mis
; NAME/KEY: primer_bind
; LOCATION: 12348..12366
; OTHER INFORMATION: 17-42-319.mis complement
; NAME/KEY: primer_bind
; LOCATION: 15222..15240
; OTHER INFORMATION: 17-41-250.mis
; NAME/KEY: primer_bind
; LOCATION: 15242..15260
; OTHER INFORMATION: 17-41-250.mis complement
; NAME/KEY: primer_bind
; LOCATION: 42199..42217
; OTHER INFORMATION: 20-841-149.mis
; NAME/KEY: primer_bind
; LOCATION: 42219..42237
; OTHER INFORMATION: 20-841-149.mis complement
; NAME/KEY: primer_bind
; LOCATION: 45423..45441
; OTHER INFORMATION: 20-842-115.mis
; NAME/KEY: primer_bind
; LOCATION: 45443..45461
; OTHER INFORMATION: 20-842-115.mis complement
; NAME/KEY: primer_bind
; LOCATION: 77039..77057
; OTHER INFORMATION: 20-853-415.mis
; NAME/KEY: primer_bind
; LOCATION: 77059..77077
; OTHER INFORMATION: 20-853-415.mis complement
; NAME/KEY: misc_binding
; LOCATION: 1227..1251
; OTHER INFORMATION: 20-828-311.probe
; NAME/KEY: misc_binding
; LOCATION: 12335..12359
; OTHER INFORMATION: 17-42-319.probe
; NAME/KEY: misc_binding
; LOCATION: 15229..15253
; OTHER INFORMATION: 17-41-250.probe
; NAME/KEY: misc_binding
; LOCATION: 42206..42230
; OTHER INFORMATION: 20-841-149.probe
; NAME/KEY: misc_binding
; LOCATION: 45430..45454
; OTHER INFORMATION: 20-842-115.probe
; NAME/KEY: misc_binding
; LOCATION: 77046..77070
; OTHER INFORMATION: 20-853-415.probe
; US-09-751-877-1

```

```

Query Match      12.0%; Score 30; DB 10; Length 81001;
Best Local Similarity 61.5%; Pred. No. 15;
Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```

```

QY 1 GGGCCCCGGCTCCAGTGGCTTCTGCTGAGTGAACACATAACAGAGCGCGGAA 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 17479 GGGCTGCTGCTCTCTCTCTCTCTGCGCGCTGATGGCCGGAACAGTATCAGGGCGGCGG 17420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 CGGGCGGGGAGGAGGGA 78
    ||||| ||||| ||||| |||||
DB 17419 CGGGCGGGGCGGCGGA 17402
    ||||| ||||| ||||| |||||

```

```

RESULT 9
US-09-867-701-8750/c
; Sequence 8750; Application US/09867701
; Patent No. US2002013237A;
; GENERAL INFORMATION:

```



```

Db 1465 TAGCATGGAGGGGCTCTGTCAGCTGAATGCTCATGTCAGGTGGAGTCTTGGAAAT 1524
Qy 192 ATGATGGAGCTGGAGGTGGGAAGAGAGTAGTAGCTTGGGCGAGCTCTCTCARGCCACCTCA 251
Db 1525 ATGATGGAGCTGGAGGTGGGAAGAGAGTAGTAGCTTGGGCGAGCTCTCTCARGCCACCTCA 1584
Qy 252 TCTGGCCAAACATCAGTCAAACTGTGAAGAGTCTAAATGATCTGCGCCCTCAAGGT 311
Db 1585 TCTGGCCAAACATCAGTCAAACTGTGAAGAGTCTAAATGATCTGCGCCCTCAAGGT 1644
Qy 312 GGCTACAAAGATATCTTTGTCAAGGTAGAGAGACCTTGTGGCTCCACGTGCACCTCCAGG 371
Db 1645 GGCTACAAAGATATCTTTGTCAAGGTAGAGAGACCTTGTGGCTCCACGTGCACCTCCAGG 1704
Qy 372 GCGTGGCTT-GCGCTCTCTACGGTCTGTCTGAGTCTCTGATGAATCTCCCTTCAGGGC 430
Db 1705 GCGTGGCTTGGGCTCTCTACGGTCTGTCTGAGTCTCTGATGAATCTCCCTTCAGGGC 1761
Qy 431 AGATTCATATTAGACTCTTCCACAGTTTGACCTGAGTTTGGCCAGATAAGGTGACATT 490
Db 1762 AGATTCATATTAGACTCTTCCACAGTTTGACCTGAGTTTGGCCAGATAAGGTGACATT 1821
Qy 491 TAGTCTGTGCTTCATGAATAGCTTAATAATTAGACATATGCTGTAGGCTTGCATT 550
Db 1822 TAGTCTGTGCTTCATGAATAGCTTAATAATTAGACATATGCTGTAGGCTTGCATT 1879
Qy 551 CCTACTCTGCTTTTCTTTCGCCCTCCAGTGTTCGGTAGTCTTTCGCCCTCCAGT 610
Db 1880 CCTACTCTGCTTTTCTTTCGCCCTCCAGTGTTCGGTAGTCTTTCGCCCTCCAGT 1938
Qy 611 CCAAGGCAACACAGATAAGTTGGAGGTGGAGTGGCTTACATAATTTTACAGACTGAA 670
Db 1939 CCAAGGCAACACAGATAAGTTGGAGGTGGAGTGGCTTACATAATTTTACAGACTGAA 1998
Qy 671 TTTCTGCTGCATCTGCATATGATATACAACTAAATACAAAGTCCCTGTGTTTATGAC 730
Db 1999 TTTCTGCTGCATCTGCATATGATATACAACTAAATACAAAGTCCCTGTGTTTATGAC 2058
Qy 731 AGGAGGCTGATCAATATAATGAATATAAGGGGGCTGGTCCCATATGCTGTGTT 790
Db 2059 AGGAGGCTGATCAATATAATGAATATAAGGGGGCTGGTCCCATATGCTGTGTT 2117
Qy 791 TTG-----TTTGTGTTCTTCTTTTGTGTTTGTGCTTCCCTCCCTCAATTA 841
Db 2118 TTGTTTGTGTTTGTGTTCTTCTTTTGTGTTTGTGCTTCCCTCCCTCAATTA 2177
Qy 842 TGAAGAGAGCAGTAAGATGTTCTCTCGGCTCTCTGAGGAGCTGGGAGCTCAGGCT 901
Db 2178 TGAAGAGAGCAGTAAGATGTTCTCTCGGCTCTCTGAGGAGCTGGGAGCTCAGGCT 2237
Qy 902 GGAATCTCCAGGAGTAGTCCGCTATCAAAATCAAAAGTCCAGGTTTGTGGGGGAA 961
Db 2238 GGAATCTCCAGGAGTAGTCCGCTATCAAAATCAAAAGTCCAGGTTTGTGGGGGAA 2297
Qy 962 AACAAAGCAGCCATACCCAGAGGACTGTCCGCTTCCCTCACCCAGCTAGGCT 1021
Db 2298 AACAAAGCAGCCATACCCAGAGGACTGTCCGCTTCCCTCACCCAGCTAGGCT 2357
Qy 1022 TTGAAGGAAACAAAGAGCAAGAAATGATTTGGCTCTGAGGAGATTCAGCTAGA 1081
Db 2358 TTGAAGGAAACAAAGAGCAAGAAATGATTTGGCTCTGAGGAGATTCAGCTAGA 2417
Qy 1082 GCTCTCTCTCCCAATCCCTCCCTCCGGCTGAGGAAACTAACAAAGGAAACAAAATTTG 1141
Db 2418 GCTCTCTCTCCCAATCCCTCCCTCCGGCTGAGGAAACTAACAAAGGAAACAAAATTTG 2477
Qy 1142 CGGAAGCAGGATTTAGAGAGAGCAAAATTCACCTGCTGCCCTTGGCTCCCGGAGCTGG 1201
Db 2478 CGGAAGCAGGATTTAGAGAGAGCAAAATTCACCTGCTGCCCTTGGCTCCCGGAGCTGG 2537
Qy 1202 ACTAGAGTCTTGGGCGCAGCCCGAGCCAGGCTTCCCGCGGCTTAGGCGCGG 1261

```

```

Db 2538 ACTAGAGAGTCTGCGGCGCAGCCCGAGCCGAGCCGCTTCCGCGGCTTAGCGCGGGG 2597
Qy 1262 GCGGCGGCGGGGAGGGGACGACAGACCGCGGACCCCTAAGACACCTGCTGTACCTCCAC 1321
Db 2598 GCGGCGGCGGGGAGGGGACGACAGACCGCGGACCCCTAAGACACCTGCTGTACCTCCAC 2657
Qy 1322 ----CCCCACCCACCCACCTCCCTCCCACTCCCTAGATGCTGCTGGGCGGCTGAACG 1376
Db 2658 CCCACACCCACCCACCCACCTCCCTCCCACTCCCTAGATGCTGCTGGGCGGCTGAACG 2717
Qy 1377 TCGCCCGCTTAAAGGGGCGGCGGCTCCACGTCTTCTGCTGAGTGAAGTGAAGTGAAC 1436
Db 2718 TCGCCCGCTTAAAGGGGCGGCGGCTCCACGTCTTCTGCTGAGTGAAGTGAAGTGAAC 2777
Qy 1437 TAAACAGAGGCGCGGAACGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1496
Db 2778 TAAACAGAGGCGCGGAACGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2837
Qy 1497 CCTCTGCGCTCGGTCCAGCCGAATCTATAAAGGAA 1532
Db 2838 CCTCTGCGCTCGGTCCAGCCGAATCTATAAAGGAA 2873

RESULT 2
US-09-846-456-1
: Sequence 1, Application US/09846456
: Patent No. US20020146792A1
: GENERAL INFORMATION:
: APPLICANT: Rosier, Marie
: APPLICANT: Prades, Catherine
: APPLICANT: Lemoine, Cendrite
: APPLICANT: Naudin, Laurent
: APPLICANT: Denefle, Patrice
: APPLICANT: Duverger, Nicolas
: APPLICANT: Brewer, Brian
: APPLICANT: Remaley, Alan
: APPLICANT: Fojo, Silvia
: TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying
: FILE REFERENCE: 3306.0505
: CURRENT APPLICATION NUMBER: US/09/846.456
: CURRENT FILING DATE: 2001-05-02
: PRIOR APPLICATION NUMBER: US 60/201,280
: PRIOR FILING DATE: 2000-05-02
: NUMBER OF SEQ ID NOS: 20
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 1
: LENGTH: 3231
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-846-456-1

Query Match      89.7%; Score 1374.4; DB 10; Length 3231;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 1504; Conservative 0; Mismatches 6; Indels 26; Gaps 10;

Qy 16 GGCCTCCACATGCATCTCCAGGCGCTGTTGG--CTCTCTATGCTGTGCTCCTGAGTGT 73
Db 1345 GCGTCCACATGCATCTCCAGGCGCTGCTTGGGCGCTCTCTATGCTCTGCTCCTGAGTGT 1404
Qy 74 CATAGAACCATCTGATGTAGTACCTGGCTTGAG--CGTGGCTGGAGATCTGTGTGACTG 132
Db 1405 GATAGAACCATCTGATGTAGTACCTGGCTTGAGCGCTGGCTGGAGATCTGTGTGACTG 1464
Qy 133 TAGCATGGAGGGGCTGT-CAGCTGATGCTGTATGTCAGGTGGTGGGAGTCTCGGAAT 191
Db 1465 TAGCATGGAGGGGCTGTGTCAGCTGAAATGTCTGATGTCAGGTGGTGGGAGTCTCGGAAT 1524
Qy 192 ATGATGGAGCTGGAGGTGGGAAGAGAGTAGGCTTGGGCGAGCTCTCTCATGCCACCTCA 251
Db 1525 ATGATGGAGCTGGAGGTGGGAAGAGAGTAGGCTTGGGCGAGCTCTCTCATGCCACCTCA 1584
Qy 252 TCTGGCCAAACATCAGTCAAACTGTGAAGAGTCTAAATGTGAATCTGCGCTTCAAGGT 311

```

```
Db 1585 TCTGCGCAAACTCAGTCAAACTGTGAAGAGCTTAAATGTGAATGTGCGCTTCAAGGT 1644
Qy 312 GGCTACAAAGATATCTGTCAAGGTAGGAGACTTGTGCGCTCCAGGTGCACTTCCAGG 371
Db 1645 GGCTACAAAGATATCTGTCAAGGTAGGAGACTTGTGCGCTCCAGGTGCACTTCCAGG 1704
Qy 372 GCGTGCCTTGGCGGCTCTAGGCTGTGCTGAGTCTTCTATGAATCTCCCTTCAGGCG 430
Db 1705 GCGTGCCTTGGCGGCTCTAGGCTGTGCTGAGTCTTCTATGAATCTCCCTTCAGGCG 1761
Qy 431 AGATTCATATTAGACTCTTACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 490
Db 1762 AGATTCATATTAGACTCTTACAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1821
Qy 490 TAGTGTGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 550
Db 1822 TAGTGTGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 1879
Qy 551 CCTACTCTGCTTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 610
Db 1880 CCTACTCTGCTTTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1938
Qy 611 CCAAGGCAAAACAGATAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 670
Db 1938 CCAAGGCAAAACAGATAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1998
Qy 671 TCTCTGCTGCTTCTACAAATGTATACAAATGTATACAAATGTATACAAATGTATACAA 730
Db 1999 TCTCTGCTGCTTCTACAAATGTATACAAATGTATACAAATGTATACAAATGTATACAA 2058
Qy 731 AGGAGGCTGATCAATATATGAATTAATGAATTAATGAATTAATGAATTAATGAATTA 790
Db 2059 AGGAGGCTGATCAATATATGAATTAATGAATTAATGAATTAATGAATTAATGAATTA 2117
Qy 791 TCG-----TTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 841
Db 2118 TTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 2177
Qy 842 TGAAGAGAGAGTAAAGATGTCTCTCGGCTCTCTGAGGAGCTGAGGAGCTGAGGCT 901
Db 2178 TGAAGAGAGAGTAAAGATGTCTCTCGGCTCTCTGAGGAGCTGAGGAGCTGAGGCT 2237
Qy 902 GGGAACTCCAGGAGTGTGCTGCTTATCAAAATCAAAATCAAAATCAAAATCAAAATCAAA 961
Db 2238 GGGAACTCCAGGAGTGTGCTGCTTATCAAAATCAAAATCAAAATCAAAATCAAAATCAAA 2297
Qy 962 AACAAAGAGAGGAGTGTGCTGCTTATCAAAATCAAAATCAAAATCAAAATCAAAATCAAA 1021
Db 2298 AACAAAGAGAGGAGTGTGCTGCTTATCAAAATCAAAATCAAAATCAAAATCAAAATCAAA 2357
Qy 1022 TTGAAGAGAGAGTGTGCTGCTTATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAAT 1081
Db 2358 TTGAAGAGAGAGTGTGCTGCTTATCAAAATCAAAATCAAAATCAAAATCAAAATCAAAAT 2417
Qy 1082 GCTCTCTCTCCCAATCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1141
Db 2418 GCTCTCTCTCCCAATCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2477
Qy 1142 CGGAAGAGAGATTTAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1201
Db 2478 CGGAAGAGAGATTTAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2537
Qy 1202 ACTAGAGAGTCTGGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1261
Db 2538 ACTAGAGAGTCTGGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2597
Qy 1262 GCGGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1321
Db 2598 GCGGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2657
Qy 1322 -----CCCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCC 1376
```

```
Db 2658 CCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCCACCC 2717
Qy 1377 TCGCCCGTTTAAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1436
Db 2718 TCGCCCGTTTAAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 2777
Qy 1437 TAAACAGAGGCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1496
Db 2778 TAAACAGAGGCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2837
Qy 1497 CCTCTGCGCTCGGTGCGAGCCGAATCTATAAAGGAA 1532
Db 2838 CCTCTGCGCTCGGTGCGAGCCGAATCTATAAAGGAA 2873
```

```
RESULT 3
US-09-846-456-3/c
; Sequence 3, Application US/09846456
; Patent No. US20020146792A1
; GENERAL INFORMATION:
; APPLICANT: Rosier, Marie
; APPLICANT: Prades, Catherine
; APPLICANT: Lemoine, Cendrine
; APPLICANT: Naudin, Laurent
; APPLICANT: Deneffe, Patrice
; APPLICANT: Duverger, Nicolas
; APPLICANT: Brewer, Bryan
; APPLICANT: Remaley, Alan
; APPLICANT: Fojo, Silvia
; TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying
; TITLE OF INVENTION: Activity and Therapeutic Uses
; FILE REFERENCE: 3806.0505
; CURRENT APPLICATION NUMBER: US/09/846.456
; CURRENT FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/201,280
; PRIOR FILING DATE: 2000-05-02
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 2893
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-846-456-3
```

```
Query Match 6.7%; Score 102.2; DB 10; Length 2893;
Best Local Similarity 65.9%; Pred. No. 2e-17;
Matches 164; Conservative 0; Mismatches 83; Indels 2; Gaps 1;

Qy 241 ATGCCACCTCATTCTGCCAAAACCTCAGGTCAAACTGTGAAGAGTCTAAATGTGAATCIG 300
Db 1820 ATGTACCTTATTCTGCCAAAACCTCAGGTCAAACTGTGAAGAGTCTAAATGTGAATCIG 1761
Qy 301 CCTTCAAGGTGGGTACAAAGGTATCTTTGTCAAGGTAGGAGACCTTTGGCCCTCCACGT 360
Db 1760 CCTTGAAGATTCATAGAAGACTCAGGACACACCGGTAGAAGAGGCCCAACGCGCCCTG 1701
Qy 361 GCACCTCCAGGCGCTGCTTGGCCCTTCTACGGGTCTCTGAGTCTTCTATGAATCIG 420
Db 1700 GAATGTACGCTGAGGCG--CACAAAGTCTCTACTCTGACAAAGATACCTTTGAGCCAC 1643
Qy 421 CCTTCAGGCGAGATTCATATTTAGACTCTTACAGTTTACCTGAGTTTGGCCAGATA 480
Db 1642 CTGAAGGCGAGATTCATATTTAGACTCTTACAGTTTACCTGAGTTTGGCCAGATA 1583
Qy 481 AGGTGACAT 489
Db 1582 AGGTGGCAT 1574
```

```
RESULT 4
US-09-846-456-1/c
; Sequence 1, Application US/09846456
; Patent No. US20020146792A1
```



```
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8332
; LENGTH: 366
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (252)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (255)
; OTHER INFORMATION: n=A,T,C or G
US-09-796-692-8332

Query Match
Best Local Similarity 2.8%; Score 43; DB 9; Length 366;
Matches 61; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 742 TCAATATAATGAATTAAGGGGCTGTCCTCATATTTCTGTTGTTTGTGTTG 801
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 264 TAAAGAAACANCAAGAGAGAGCTGATCCCAAGCTACAGGGTTTTCGTTG 205
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 802 TTTCCTTTTCTGTTTGTGGCTCCTTCCTCT 834
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 204 TTGTTGTTGTTGTTGGAGAGTCTGCTCT 172

RESULT 8
US-09-764-855-50/C
; Sequence 60, Application US/09764855
; Patent No. US2002011991A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P110
; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
; LENGTH: 251
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-855-60

Query Match
Best Local Similarity 2.7%; Score 41.8; DB 10; Length 251;
Matches 61; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 772 CCATATGTTCTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 831
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 248 CCGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 189
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 832 TCTCAATTATGAGAGAGAGTGAAGATGTC 864
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188 TTCAATTTTCAATAAATTAATTAATCTGTAC 156

RESULT 9
US-09-834-975-994
; Sequence 994, Application US/09834975
; Patent No. US20020110815A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Brown, Jeffrey
; APPLICANT: Bolt, Andrew

; APPLICANT: Van Huffel, Christophe
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
; TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-016B
; CURRENT APPLICATION NUMBER: US/09/834,975
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,538
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 1046
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 994
; LENGTH: 1974
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-834-975-994

Query Match
Best Local Similarity 2.7%; Score 41.6; DB 10; Length 1974;
Matches 65; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 717 TGTGTTTATCACAGGAGGCTGATCAATATAATTAAGGGGCTGTCCTCAT 776
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 221 TTGTTTGGGGGGGGGGGGGTTTATTAAGGGGCCCCCCCCCT 280
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 777 ATGCTCTCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 820
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 281 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 324

RESULT 10
US-09-764-870-14/C
; Sequence 14, Application US/09764870
; Patent No. US20020042386A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT214
; CURRENT APPLICATION NUMBER: US/09/764,870
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 646
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 2099
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-870-14

Query Match
Best Local Similarity 2.7%; Score 41.4; DB 10; Length 2099;
Matches 57; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 776 TATGTTCTGCTTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 835
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2094 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2035
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 836 AATTATCAAGACAGCAGTAAG 858
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2034 CTTTATAAAACTAGAGAIAAG 2012

RESULT 11
US-09-764-877-3972
; Sequence 3972, Application US/09764877
; Patent No. US2002014740A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
```

; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3972
; LENGTH: 12718
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3972

Query Match 2.7%; Score 41; DB 10; Length 12718;

Best Local Similarity 60.2%; Pred. No. 1.3;
Matches 68; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 730 CAGGAGGTGATCAATAAATCAATTAAGGGGCTGGTCCCAATATGTCCTGTT 789

DB 7756 CTGGGTGACAGACAGACCCCTGCTCTTAAAGAAATGCAACATATATCTTTT 7815

QY 790 TTGTTGTTTCTTTCTTTTCTTTTCTGCTCTCTCTCAATTAT 842

DB 7816 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 7868

RESULT 12

US-09-878-574-4299

; Sequence 4299, Application US/09878574

; Patent No. US20020110548A1

; GENERAL INFORMATION:

; APPLICANT: Byrium, Joseph R.

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Thompson, Michael D.

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with:

; FILE OF INVENTION: Plants

; FILE REFERENCE: 38-21(1540)B

; CURRENT APPLICATION NUMBER: US/09/878,574

; CURRENT FILING DATE: 2003-12-21

; PRIOR APPLICATION NUMBER: 09/333,535

; PRIOR FILING DATE: 1999-06-14

; NUMBER OF SEQ ID NOS: 15775

; SEQ ID NO 4299

; LENGTH: 545

; TYPE: DNA

; ORGANISM: Glycine max

; FEATURES:

; NAME/KEY: unsure

; LOCATION: (1)..(545)

; OTHER INFORMATION: unsure at all n locations

; OTHER INFORMATION: Clone ID: LIB3028-013-Q1-B1-H1

US-09-878-574-4299

Query Match

Best Local Similarity 62.1%; Score 40.6; DB 10; Length 545;

Matches 64; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 776 TATGTCGCTCTTTGTTTCTTTCTTTCTTTTCTTTTCTTTTCTTTCTCTC 835

DB 266 TTTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 325

QY 836 AATTATGAAGAGCAGTAAGATGTCCTCTCGGTCCTCT 878

DB 326 CCTTTTCT 368

RESULT 13

US-10-008-118A-19/c

; Sequence 19, Application US/10008-18A

; Publication No. US20020187539A1

; GENERAL INFORMATION:

; APPLICANT: Cahoon, Rebecca E.

; APPLICANT: Weng, Zude

; TITLE OF INVENTION: Plant MYB-Related Transcription Factors

; FILE REFERENCE: BB1280 USDI

; CURRENT APPLICATION NUMBER: US/10/008,118A

; CURRENT FILING DATE: 2001-12-05

; PRIOR APPLICATION NUMBER: 60/109,294

; PRIOR FILING DATE: 1998-11-20

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 19

; LENGTH: 1372

; TYPE: DNA

; ORGANISM: Zea mays

US-10-008-118A-19

Query Match

Best Local Similarity 60.6%; Score 40.2; DB 9; Length 1372;

Matches 66; Conservative 3; Mismatches 43; Indels 0; Gaps 0;

QY 776 TATGTCGCTCTTTGTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTCTC 835

DB 1341 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1282

QY 836 AATTATGAAGAGCAGTAAGATGTCCTCTCGGTCCTCTGAGGA 884

DB 1281 TTTTCTTCCAGAGACCCAAAAGATTTCGCCCATGATCATTTGGCGA 1233

RESULT 14

US-09-443-704-19/c

; Sequence 19, Application US/09443704

; Patent No. US20020066120A1

; GENERAL INFORMATION:

; APPLICANT: Cahoon, Rebecca E.

; APPLICANT: Liu, Zhan-Bin

; APPLICANT: Odell, Joan

; APPLICANT: Rafalski, Antoni

; APPLICANT: Shi, June

; APPLICANT: Weng, Zude

; TITLE OF INVENTION: Plant MYB-Related Transcription Factors

; FILE REFERENCE: BB1280 US NA

; CURRENT APPLICATION NUMBER: US/09/443,704

; CURRENT FILING DATE: 1999-11-19

; EARLIER APPLICATION NUMBER: 60/109,294

; EARLIER FILING DATE: No. US20020066120A1ember 20, 1998

; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 19

; LENGTH: 1372

; TYPE: DNA

; ORGANISM: Zea mays

US-09-443-704-19

Query Match

Best Local Similarity 60.6%; Score 40.2; DB 10; Length 1372;

Matches 66; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 776 TATGTCGCTCTTTGTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTCTC 835

DB 1341 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1282

QY 836 AATTATGAAGAGCAGTAAGATGTCCTCTCGGTCCTCTGAGGA 884

DB 1281 TTTTCTTCCAGAGACCCAAAAGATTTCGCCCATGATCATTTGGCGA 1233

RESULT 15

US-09-764-869-2064/c

; Sequence 2064, Application US/09764869

; Patent No. US20020061521A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC007

; CURRENT APPLICATION NUMBER: US/09/764,869

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper

; NUMBER OF SEQ ID NOS: 2442

; SOFTWARE: PatentIn Ver. 2.0

```
; SEQ ID NO 2064
; LENGTH: 16511
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (14362)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-869-2064

Query Match      2.68; Score 40; DB 10; Length 16511;
Best Local Similarity 51.78; Pred.No. 2.9;
Matches 91; Conservative 0; Mismatches 85; Indels 0; Gaps 0:

QY 580 TGCACCTTCACAAATGTATACAAACTAATATACAGTCCTGTGTTTTTATCATCAGAGGAGGCT 739
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 1103 TGTACTTTTTCGATAAATTCCTGAGAAATTTTTCCTCAIGTTTATGGCCATAGGTAGTTT 1044
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 740 GATCAATATAATGAATTAAGAGGGGCTGTCCCATATCTCTGTCTGTTTGTGTTGTT 799
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DQ 1043 TTGTCATTTTAACTACAAATGTTCATTTTATTTTTCCTCTTTTATTTTCA 984
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 800 TGTTCCTTTTGTGTTTGTGGCCCTCTCTCAATTTATGAAGAGAGAGCACT 855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 983 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 928
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Search completed: February 3, 2003, 16:28:23
Job time : 365.471 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

R-un on: February 3, 2003, 16:38:05 ; Search time 1614.75 Seconds
(without alignments)
10165.006 Million cell updates/sec

Title: US-09-596-141c-3_COPY_1080_1643

Perfect score: 564

Sequence: 1 gagctctctctccccaatc.....cagggcaggcgaggagctc 564

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_hg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pi.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_hg_hum.*

31: em_hg_inv.*

32: em_hg_other.*

33: em_hg_mus.*

34: em_hg_pln.*

35: em_hg_rod.*

36: em_hg_sam.*

37: em_hg_vrt.*

38: em_sy.*

39: em_hgo_hum.*

40: em_hgo_mus.*

41: em_hgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	564	100.0	1167	9	HSA252201	AT252201 Homo sapi
2	564	100.0	1643	6	AX060715	AX060715 Sequence
3	564	100.0	1643	6	AX060894	AX060894 Sequence
c	564	100.0	96717	9	AL359182	AL359182 Human DNA
	564	100.0	175064	2	AC012230	AC012230 Homo sapi
5	564	100.0	175064	2	AC012230	AC012230 Homo sapi
6	560.8	99.4	1167	9	AP258623	AP258623 Homo sapi
7	548	97.2	183999	6	AX092589	AX092589 Sequence
8	547.4	97.1	3231	6	AX351029	AX351029 Sequence
9	547.4	97.1	149034	9	AF275948	AF275948 Homo sapi
10	541	95.9	201144	9	AF287262	AF287262 Homo sapi
11	519.8	92.2	69570	2	AC021246	AC021246 Homo sapi
12	456.4	80.9	2893	6	AX351031	AX351031 Sequence
13	271.6	48.2	69570	2	AC021246	AC021246 Homo sapi
c	229.8	40.7	90698	2	AC021345	AC021345 Homo sapi
15	173	30.7	697	9	AF258627	AF258627 Homo sapi
c	155.4	27.6	186889	2	AL080724	AL080724 Mus muscu
17	153.2	27.2	278572	10	AF287263	AF287263 Mus muscu
c	153.2	27.2	145833	2	AC125837	AC125837 Rattus no
19	98	17.4	7260	6	AX253452	AX253452 Sequence
20	91	16.1	221	6	AX351032	AX351032 Sequence
21	89.4	15.9	1556	9	AK024328	AK024328 Homo sapi
22	89	15.8	1750	9	AK022254	AK022254 Homo sapi
23	75	13.3	9854	6	AX127831	AX127831 Sequence
24	75	13.3	9854	6	AX139818	AX139818 Sequence
25	67	11.9	10442	6	AX060713	AX060713 Sequence
26	67	11.9	10442	6	AX060892	AX060892 Sequence
27	67	11.9	10442	9	AF285167	AF285167 Homo sapi
28	60	10.6	10474	6	AX060719	AX060719 Sequence
29	60	10.6	10474	6	AX060721	AX060721 Sequence
30	60	10.6	10474	6	AX060898	AX060898 Sequence
31	60	10.6	10474	6	AX060900	AX060900 Sequence
32	55.6	9.9	191720	2	AC109982	AC109982 Rattus no
c	50.4	8.9	125020	9	AF429315	AF429315 Homo sapi
34	49.6	8.8	96136	2	AC097674	AC097674 Rattus no
35	49.6	8.8	125020	9	AF429315	AF429315 Homo sapi
c	49.4	8.8	173127	2	AC131142	AC131142 Rattus no
37	49	8.7	65780	2	AC111741	AC111741 Rattus no
c	47.4	8.4	180461	2	AC106563	AC106563 Rattus no
39	47	8.3	110000	2	AC096324	AC096324 Rattus no
c	47	8.3	167886	2	AC126316	AC126316 Rattus no
41	46.6	8.3	55357	2	AC121553	AC121553 Mus muscu
c	46.4	8.2	184402	2	AC127041	AC127041 Rattus no
c	46.2	8.2	150305	2	AC114705	AC114705 Rattus no
c	46.2	8.2	209216	2	AC117126	AC117126 Rattus no
c	46	8.2	97967	2	AC126984	AC126984 Rattus no

ALIGNMENTS

RESULT 1
HSA252201
LOCUS
DEFINITION
Homo sapiens partial ABC-1 gene for ATP-binding cassette transporter-1, 5'UTR and promoter region.
ACCESSION
AJ252201
VERSION
ABC-1 gene; GI:12053757
KEYWORDS
ABC-1 gene; ATP-binding cassette transporter-1; promoter.
SOURCE
Human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1167)
Porsch-Ozcurumez.W., Langmann.T., Heimerl.S., Borsukova.H.,
Kaminski.W.E., Drobniak.W., Honer.C., Schumacher.C. and Schmitz.G.

Pred. No. is the number of results predicted by chance to have a

chr	start	end	feature	source
*	48117	48216	gap of 100 bp	
*	48217	52618	contig of 4402 bp	
*	52619	52718	gap of 100 bp	
*	52719	56592	contig of 3874 bp	
*	56593	56692	gap of 100 bp	
*	56693	59633	contig of 2943 bp	
*	59636	59735	gap of 100 bp	
*	59736	63661	contig of 3926 bp	
*	63662	63761	gap of 100 bp	
*	63762	68437	contig of 4576 bp	
*	68438	68537	gap of 100 bp	
*	68538	71458	contig of 2921 bp	
*	71459	71558	gap of 100 bp	
*	71559	76888	contig of 5330 bp	
*	76889	76988	gap of 100 bp	
*	76989	82113	contig of 5125 bp	
*	82114	82213	gap of 100 bp	
*	82214	88220	contig of 6007 bp	
*	88221	89320	gap of 100 bp	
*	89321	93499	contig of 5179 bp	
*	93500	93599	gap of 100 bp	
*	93600	97901	contig of 4302 bp	
*	97902	98001	gap of 100 bp	
*	98002	103016	contig of 5015 bp	
*	103017	103116	gap of 100 bp	
*	103117	109178	contig of 6062 bp	
*	109179	109278	gap of 100 bp	
*	109279	127307	contig of 8029 bp	
*	127308	127407	gap of 100 bp	
*	127408	124079	contig of 6672 bp	
*	124080	124179	gap of 100 bp	
*	124180	131281	contig of 7102 bp	
*	131282	131381	gap of 100 bp	
*	131382	138059	contig of 6678 bp	
*	138060	138159	gap of 100 bp	
*	138160	145491	contig of 7332 bp	
*	145492	145591	gap of 100 bp	
*	145592	157391	contig of 11800 bp	
*	157392	157491	gap of 100 bp	
*	157492	175064	contig of 17573 bp	
RES Location/Qualifiers				
source				
1..175064				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/clone="RP11-1M13"				
/clone_lib="RP11 Human Male BAC"				
1..1003				
misc_feature				
/note="assembly_fragment"				
104..2634				
misc_feature				
/note="assembly_fragment"				
2735..4415				
misc_feature				
/note="assembly_fragment"				
4516..5785				
misc_feature				
/note="assembly_fragment"				
5886..7879				
misc_feature				
/note="assembly_fragment"				
7980..9686				
misc_feature				
/note="assembly_fragment"				
9787..12253				
misc_feature				
/note="assembly_fragment"				
12354..15228				
misc_feature				
/note="assembly_fragment"				
15329..17200				
misc_feature				
/note="assembly_fragment"				
17301..20131				
misc_feature				
/note="assembly_fragment"				
20232..22587</				

[illegible]

9. RESULT


```
repeat_region /rpt_family="Alu" 5563. .5839
repeat_region /rpt_family="Alu" 6615. .6877
repeat_region /rpt_family="Alu" 8800. .9046
repeat_region /rpt_family="Alu" 9998. .10279
repeat_region /rpt_family="Alu" 11859. .12133
repeat_region /rpt_family="Alu" 12810. .12902
repeat_region /rpt_family="Alu" 15220. .15399
repeat_region /rpt_family="Alu" 15508. .15787
repeat_region /rpt_family="Alu" 18600. .18860
repeat_region /rpt_family="Alu" 20950. .21206
repeat_region /rpt_family="Alu" 25036. .25314
repeat_region /rpt_family="Alu" 25036. .25314
CDS
join(25924. .25989,40385. .40478,45012. .45153,46423. .46541,
67729. .67850,70831. .71007,83960. .84052,89010. .89250,
91962. .92101,92433. .92549,95758. .96955,97702. .97907,
98428. .98604,100391. .100613,102360. .102581,103642. .103846,
104951. .105064,106862. .107033,108023. .108154,
109460. .109602,109806. .109943,110646. .110866,
112424. .112196,113183. .113385,115053. .115201,
115297. .115410,116806. .116954,118604. .118728,
119964. .120062,123094. .123283,124804. .124898,
126208. .126240,127363. .127458,128943. .129017,
129539. .129708,130939. .131116,133114. .133229,
133341. .133485,134326. .134649,135736. .135865,
136130. .136250,137038. .137200,138008. .138114,
140469. .140610,140982. .141216,142080. .142163,
142646. .142738,143397. .143640,144581. .144721)
/gene="ABCA1"
/codon_start=1
/product="ABCA1"
/protein_id="AAF96276.1"
/db_xref="GI:9247086"
/translation="MACPOLRLLLWKNLTFRRRTQCLLEVAWPFLFILLSVRL
SYYPEGHEHPKAMPFSAGTLPWVGIIICNANPCFRTPGAPGVGNFNKSIY
ARLPSDARRLLYSOKDTSMKDMKVLRTLQIKKSSNLKLODPLVDNFTSGFLYH
NLSLPASTVDKMLRADVLHKVPLQGVOLHTSLCNGSKSEEMIQLDQVSEICGLP
BEKLAARVLASNMWILKPIRLINSTSPFPPSKELAEAKILLHSLGTLAQELFSR
SWDMRQEVFLTNVNSSSTQI:QAVSRIVCGPPEGGGKIKSLNWIENNNYKALF
GGNTEDEDAETFDNSTPYCNDLKNLLESSPLSRIIWKALKPLLVGKILYPTPAT
QQLDLQWTADQIVAFIAKHDPEDVSSNGSVYTWREAFNETQNAIRTISRMECVNLN
RQVMAEYKTFQELAVFHDLEGWEEELSPKIWTPEMSQEMDVRLMLDSRDNDHFW
KLEPIATEVWLIKNSMELLDERKFWAGIVFTGTPGSLIEPHVVKYAIRMDIDNERT
NKIDGYDQCPRADPREDMRYVWGFEAYLODQVVEQALIRVLIGTEKKIGYMQMPY
PCYVDDIPLRVMSMSMLTMLAWIYSAVVIKGIIVIEKARKLRETRIMGLDNLNLW
FSWFISSILPILVAGGLLVLKLGNNLLPYSDPSVEVFLSVFAVTLIQCLFLSTLP
SRNALAACGGIITYTLYPVLCVAMQYVFTKIFAXLLSPVAFGEGCEYFALFE
EQGIQVQNDNLFPESPEDEGNLTYSMMFLFTFLYCVMTWYIEAVPFGYGIPLRPW
YFPCYVFAKFGESDEKSHPGSNOKRISIEICMEEPHLLKGVSIONLAKVYRDMKV
AVDGLALNFYEGQITSFLGNGAGKTTMTSLTGLFPTSTAVILGKDIISEMSIR
QNLGVQPHNVLFMLAVBEHINWYALKGLSEKHVKAEMEOMALDVLPLPSKLSKT
SOLSGWMOKLUSALAFVGGSKVILDEPTAGVDYFYSRRGIWELLKYRQCTIILT
HMDPEADVLGRIAIISHGKCCVGSLLFLKNQGTGYTLVLKRDVSSSSRGNSS
STVYLKEDSDVSGSDAGLGDHSDIETLIDVSAISNLIRKHSFARLVEDIGHCL
TYVLPYAAKGAAGVFLPHEIDRLDGLISYSGISSEPTLEEILPKVAEPSSVDAETS
DGLTPARNRRAFGCKQSCIRPPTEDDAANDSDIDPESRETLLSGMDKGYQVK
GWLKIQOQFVALLKRLGRIARRSRGFFAQIVLPAVFCVIALVFSLLVPPGKYPSE
LQPMWYEQYTFVNSNDAPETGTLELLNALTKDPCFTRCMEGNIPDTPCQAGEEW
TTAVPQTIMDLFQNGWNTMQNPSAPQCSSDKIKMLPVPGPAGGLPLPPQPKQNTA
DILQQLGRNLSVLVYKTYVQI IAKSLKNKWNVNEFRYGGFSLGVSNTQALPPQOEVA
DAXKOMKHLKLADESSADRLNSLRPMIGLDTENNKKVFNKNGKHAISSFLNLIN
NALLRANLQKGENSHGQITAFNPLNLTIKQOLSEVAXMT:SVQVLSICVIFPMSV
PASFVVFLLQENVSARKHLQFISGVKPVYIWLSEVFNMDNCNYVYPATILVIFIFCQ
```

Query Match

Best Local Similarity 97.1%; Score 547.4; DB 9; Length 149034;

Matches 563; Conservative 0; Mismatches 1; Indels 5; Gaps 1;


```
PASTVFLIOERYSKAKHLQFIQGVKVIYWLNSFWDMCNVYVPATLVILIFICFOQ
KSYVSSNVLVALLLILLYGWSHPLMYPASVFKIEPTAYVVLTSYNLZIGINGSVA
IFVLELPTCNKLNINDILKSVFLIFPHCLGLIDMKVKNQAMADALEFGENREVS
PLSWDLNRLNFAMAVEGVVFFLLITVLIQVRFIRPRVRAKLSPLNDEDEDVRRERQ
RLDGGQNDLEIKELTK-YRRKRKPAVDRCVGI PPGECLGLVNGACKSSTFRM
LTGQDVTIRGDAFENKLSILSNIEHVNMGYCPQDAITELLTGRHVEFFALLRGV
PEXSVGVGEWAIRKILGLVKGEXVAGNYSGNKRKLSTAMAL-GGPVVFLEPTTG
KDPKARPLNICALSVYKESRSVVL-SHSMEECEALCTRMAIKVNGRFRCLGSYQHLEK
NRFSGDVIIVRIAGSPDLKPVDQFFGLAFPGSVLKEKKHRMLQYQLPSSLSSARI
FSILQDEKYESYV
72893..72986
/gene="ABCA1"
/number=3
exon
77520..77661
/gene="ABCA1"
/number=4
exon
78931..79049
/gene="ABCA1"
/number=5
exon
101048..101169
/gene="ABCA1"
/number=6
exon
104132..104328
/gene="ABCA1"
/number=7
exon
117287..117379
/gene="ABCA1"
/number=8
exon
122329..122569
/gene="ABCA1"
/number=9
exon
125285..125424
/gene="ABCA1"
/number=10
exon
125754..125870
/gene="ABCA1"
/number=11
exon
130079..130276
/gene="ABCA1"
/number=12
exon
131023..131228
/gene="ABCA1"
/number=13
exon
131749..131925
/gene="ABCA1"
/number=14
exon
133708..133930
/gene="ABCA1"
/number=15
exon
135677..135898
/gene="ABCA1"
/number=16
exon
136959..137163
/gene="ABCA1"
/number=17
exon
138268..138381
/gene="ABCA1"
/number=18
exon
140179..140350
/gene="ABCA1"
/number=19
exon
141340..141471
/gene="ABCA1"
/number=20
exon
142777..142919
/gene="ABCA1"
/number=21
exon
143123..143260
/gene="ABCA1"
/number=22
exon
143963..144183
/gene="ABCA1"
```

Query Match 95.9%; Score 541; DB 9; Length 201144;

Best Local Similarity 98.2%; Preg. No. 2.1e-119; Mismatches 5; Indels 5; Gaps 1;

Matches 559; Conservative 0;

```
QY 1 GAGCTCTCTCTCCCAATCCCTCCCTCGGCTGAGGAACTAACAAAGAAAAAAT 60
DB 33453 GAGCTCTCTCTCCCAATCCCTCCCTCGGCTGAGGAACTAACAAAGAAAAAAT 33512
QY 61 TCGGGAACAGAGATTAGAGGAGCAAAATCCACTGGTGGCTTGGCTGGCGGAGCGT 120
DB 33513 TCGGGAACAGAGATTAGAGGAGCAAAATCCACTGGTGGCTTGGCTGGCGGAGCGT 33572
QY 121 GGACTAGAGAGTCTGCGGCGCAGCCCGCAGCCCGCTTCCGCGCGCTTTAGGCGGCG 180
DB 33573 GGACTAGAGAGTCTGCGGCGCAGCCCGCAGCCCGCTTCCGCGCGCTTTAGGCGGCG 33632
QY 181 GGGCCCGGGGGGGAAGGGGAGCGAGACCGCGGAGCCCTTAAGACACCTGTGTACCTCC 240
DB 33633 GGGCCCGGGGGGGAAGGGGAGCGAGACCGCGGAGCCCTTAAGACACCTGTGTACCTCC 33692
QY 241 AC-----CCCAACCCACACACCTCCCACTCCCACTAGATGTCTGCTGGGGGCGTGA 295
DB 33693 ACCCAACCCACACACCTCCCACTCCCACTAGATGTCTGCTGGGGGCGTGA 33752
QY 296 CGTCGCCGCTTTAAGGGGGGGCGCGGCTCCACGTCTTCTGCTAGTGTACTGAACTA 355
DB 33753 CGTCGCCGCTTTAAGGGGGGGCGCGGCTCCACGTCTTCTGCTAGTGTACTGAACTA 33812
QY 356 CATAACAGAGCGCGGAGACCGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 415
DB 33813 CATAACAGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 33872
QY 416 AACCTCTGCGCTCGGTGCAGCCGAATCTATAAAGAACTAGTCCCGGCAAAAACCCCGT 475
DB 33873 AACCTCTGCGCTCGGTGCAGCCGAATCTATAAAGAACTAGTCTCGGCAAAAACCCCGT 33932
QY 476 AATTGGGAGCGAGAGTGAAGTGGGGGGGACCGCGAGAGCGGAGCGGAGCGGAGCGGAGCGGAG 535
DB 33933 AATTGGGAGCGAGTGAAGTGGGGGGGACCGCGAGAGCGGAGCGGAGCGGAGCGGAGCGGAG 33992
QY 536 GGGCTGGCGAGGCGGAGGCGGGGAGGCTC 564
DB 33993 GGGCTGGCGAGGCGGAGGCGGGGAGGCTC 34021
```

RESULT 11

AC021246

LOCUS

69570 bp

DNA

linear

HTG 13-JUL-2000


```

DEFINITION Homo sapiens clone RP11-IN10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC021246
VERSION AC021246.2 GI:9119882
KEYWORDS HNG; HTGS_PHASE0.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 69570)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
TITLE Homo sapiens chromosome, clone RP11-IN10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 69570)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boquelavkiy,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
Choele,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Deaerilao,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kahn,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGuirk,A., McKernan,K.,
McPheeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A., and Zody,M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6705871.
All repeats were identified using repeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2512
Center clone name: L_N10
-----
* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 872 971: gap of 100 bp
* 972 1834: contig of 863 bp in length
* 1835 1934: gap of 100 bp
* 1935 2804: contig of 870 bp in length
* 2805 2904: gap of 100 bp
* 2905 3745: contig of 841 bp in length
* 3745 3845: gap of 100 bp
* 3845 4696: contig of 851 bp in length
* 4697 4796: gap of 100 bp
* 4797 5640: contig of 844 bp in length
* 5641 5740: gap of 100 bp
* 5741 6540: contig of 800 bp in length
* 6541 6640: gap of 100 bp
* 6641 7509: contig of 868 bp in length
* 7510 7609: gap of 100 bp
*
* 8479: contig of 870 bp in length
* 8480 8579: gap of 100 bp
* 8580 9430: contig of 851 bp in length
* 9431 9530: gap of 100 bp
* 9531 10376: contig of 846 bp in length
* 10377 10476: gap of 100 bp
* 10477 11322: contig of 846 bp in length
* 11323 11422: gap of 100 bp
* 11423 12302: contig of 880 bp in length
* 12303 12402: gap of 100 bp
* 12403 13280: contig of 878 bp in length
* 13281 13380: gap of 100 bp
* 13381 14241: contig of 861 bp in length
* 14242 14341: gap of 100 bp
* 14342 15196: contig of 855 bp in length
* 15197 15296: gap of 100 bp
* 15297 16123: contig of 827 bp in length
* 16124 16223: gap of 100 bp
* 16224 17072: contig of 849 bp in length
* 17073 17172: gap of 100 bp
* 17173 18041: contig of 869 bp in length
* 18042 18141: gap of 100 bp
* 18142 19009: contig of 868 bp in length
* 19010 19109: gap of 100 bp
* 19110 19966: contig of 857 bp in length
* 19967 20066: gap of 100 bp
* 20067 20921: contig of 855 bp in length
* 20922 21021: gap of 100 bp
* 21022 21865: contig of 844 bp in length
* 21866 21965: gap of 100 bp
* 21966 22832: contig of 867 bp in length
* 22833 22932: gap of 100 bp
* 22933 23780: contig of 848 bp in length
* 23781 23880: gap of 100 bp
* 23881 24733: contig of 853 bp in length
* 24734 24833: gap of 100 bp
* 24834 25670: contig of 837 bp in length
* 25671 25770: gap of 100 bp
* 25771 26621: contig of 851 bp in length
* 26622 26721: gap of 100 bp
* 26722 27578: contig of 855 bp in length
* 27577 27676: gap of 100 bp
* 27677 28532: contig of 856 bp in length
* 28533 28632: gap of 100 bp
* 28633 29492: contig of 860 bp in length
* 29493 29592: gap of 100 bp
* 29593 30455: contig of 863 bp in length
* 30456 30555: gap of 100 bp
* 30556 31410: contig of 855 bp in length
* 31411 31510: gap of 100 bp
* 31511 32368: contig of 858 bp in length
* 32369 32468: gap of 100 bp
* 32469 33312: contig of 844 bp in length
* 33313 33412: gap of 100 bp
* 33413 34268: contig of 856 bp in length
* 34269 34368: gap of 100 bp
* 34369 35204: contig of 836 bp in length
* 35205 35304: gap of 100 bp
* 35305 36156: contig of 852 bp in length
* 36157 36256: gap of 100 bp
* 36257 37128: contig of 872 bp in length
* 37129 37228: gap of 100 bp
* 37229 38083: contig of 855 bp in length
* 38084 38183: gap of 100 bp
* 38184 39031: contig of 848 bp in length
* 39032 39131: gap of 100 bp
* 39132 40006: contig of 875 bp in length
* 40007 40106: gap of 100 bp
* 40107 40967: contig of 861 bp in length
* 40968 41067: gap of 100 bp
* 41068 41913: contig of 846 bp in length
* 41914 42013: gap of 100 bp
* 42014 42824: contig of 811 bp in length

```

```
* 42925 42924: gap of 100 bp
* 42925 43776: contig of 852 bp in length
* 43777 43876: gap of 100 bp
* 43877 44752: contig of 876 bp in length
* 44753 44852: gap of 100 bp
* 44853 45724: contig of 872 bp in length
* 45725 45824: gap of 100 bp
* 45825 46643: contig of 819 bp in length
* 46644 46743: gap of 100 bp
* 46744 47599: contig of 856 bp in length
* 47600 47699: gap of 100 bp
* 47700 48551: contig of 852 bp in length
* 48552 48651: gap of 100 bp
* 48652 49485: contig of 834 bp in length
* 49486 49585: gap of 100 bp
* 49586 50440: contig of 855 bp in length
* 50441 50540: gap of 100 bp
* 50541 51404: contig of 864 bp in length
* 51405 51504: gap of 100 bp
* 51505 52372: contig of 868 bp in length
* 52373 52472: gap of 100 bp
* 52473 53328: contig of 856 bp in length
* 53329 53428: gap of 100 bp
* 53429 54268: contig of 840 bp in length
* 54269 54368: gap of 100 bp
* 54369 55229: contig of 851 bp in length
* 55230 55329: gap of 100 bp
* 55330 56197: contig of 858 bp in length
* 56198 56297: gap of 100 bp
* 56298 57163: contig of 866 bp in length
* 57164 57263: gap of 100 bp
* 57264 58130: contig of 867 bp in length
* 58131 58230: gap of 100 bp
* 58231 59082: contig of 852 bp in length
* 59083 59182: gap of 100 bp
* 59183 60020: contig of 838 bp in length
* 60021 60120: gap of 100 bp
* 60121 60983: contig of 863 bp in length
* 60984 61083: gap of 100 bp
* 61084 61935: contig of 852 bp in length
* 61936 62035: contig of 831 bp in length
* 62036 62966: gap of 100 bp
* 62967 63827: contig of 861 bp in length
* 63828 63927: gap of 100 bp
* 63928 64783: contig of 856 bp in length
* 64784 64883: gap of 100 bp
* 64884 65740: contig of 857 bp in length
* 65741 65840: gap of 100 bp
* 65841 66684: contig of 844 bp in length
* 66685 66784: gap of 100 bp
* 66785 67651: contig of 867 bp in length
* 67652 67751: gap of 100 bp

Query Match 92.2%; Score 519.8; DB 2; Length 69570;
Best Local Similarity 99.3%; Pred. No. 2.5e-114;
Matches 543; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 20 CCGTCCTCCGCTGAGGAACCTACAAA-GGAAAAAAAATTGCGGAAGCAGGATTTA 78
DB 41109 CCTCCCTCCGCTGAGGAACCTACAAAAGGAAAAAAAATTGCGGAAGCAGGATTTA 41168

QY 79 GAGGAAGCAAAATCCACTGGTGGCCCTGGCTGCGGGAGACGTGGACTAGAGATCTGCGG 138
DB 41169 GAGGAAGCAAAATCCACTGGTGGCCCTGGCTGCGGGAGACGTGGACTAGAGATCTGCGG 41228

QY 139 CGAGCCCGGAGCCCGAGCGCTTCCCGCGGCTTTAGCCCGGGGGGCCCGGGGGGAAG 198
DB 41229 CGAGCCCGGAGCCCGAGCGCTTCCCGCGGCTTTAGCCCGGGGGGCCCGGGGGGAAG 41288

QY 199 GGGACGACGACCCGCGACCCCTAGACACCTGCTGTACCTCCACCCCGACCCCGACCC 258
DB 41289 GGGACGACGACCCCGCGACCCCTAGACACCTGCTGTACCTCCACCCCGACCCCGACCC 41348
```

```
QY 259 TCOCCTCAACT-CCCTAGATGTGTCGTGGGGGCTGAACGTCGCCGCTTAAAGGGCGGG 317
DB 41349 TCOCCTCAACTCCCTATA-TGTCGTGGGGGCGCTGAACGTCGCCGCTTAAAGGGCGGG 41408

QY 318 CCGCGGCTCCACGTCGCTTC-TGTGAGTGAAGTGAACATATAAAGAGGCGGGGAAGG 377
DB 41409 CCGCGGCTCCACGTCGCTTC-TGTGAGTGAAGTGAACATATAAAGAGGCGGGGAAGG 41468

QY 378 GCGGGGAGGAGGAGACACAGGCTTIGACCGATAGTAAGCTTCGCTCGGTGCAGCC 437
DB 41469 GCGGGGAGGAGGAGACACAGGCTTIGACCGATAGTAAGCTTCGCTCGGTGCAGCC 41528

QY 438 GAATCTATAAAGAACTAGTCCCGGCAAAAACCCGCTTAATTCGGAGCGAGTGAAGTGG 497
DB 41529 GAATCTATAAAGAACTAGTCCCGGCAAAAACCCGCTTAATTCGGAGCGAGTGAAGTGG 41588

QY 498 GCGGGGACCGCGAGCGGAGCGGACCTTCTCTCCGGGCTGCGGAGGCGAGCGGG 557
DB 41589 GCGGGGACCGCGAGCGGAGCGGACCTTCTCTCCGGGCTGCGGAGGCGAGCGGG 41648

QY 558 GGAGCTC 564
DB 41649 GGAGCTC 41655

RESULT 12
LOCUS AX351031 2893 bp DNA linear PAT C6-FEB-2002
DEFINITION Sequence 3 from Patent WO0183746.
ACCESSION AX351031
VERSION AX351031.1 GI:18615387
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Rosier-Montus,M.F., Prades,C., Lemoine,C., Naudin,L., Deneffe,P.,
Brewer,B., Duvergier,N., Remaley,A. and Santamarina-Fojo,S.
TITLE Regulatory nucleic acid sequences of the abcl1 gene
JOURNAL Patent: WO 0183746-A 3 08-NOV-2001;
Aventis Pharma S.A. (FR)
FEATURES
source
BASE COUNT 745 a 681 c 769 g 598 t
ORIGIN
Query Match 80.9%; Score 456.4; DB 6; Length 2893;
Best Local Similarity 98.7%; Pred. No. 3.9e-99;
Matches 472; Conservative 0; Mismatches 1; Indels 5; Gaps 1;
```



```

34369 35204: contig of 836 bp in length
35205 35304: gap of 100 bp in length
35305 35156: contig of 852 bp in length
36257 36256: gap of 100 bp in length
37129 37128: contig of 872 bp in length
37229 36883: contig of 855 bp in length
38084 38183: gap of 100 bp in length
38184 39031: contig of 848 bp in length
39032 39131: gap of 100 bp in length
39132 40006: contig of 875 bp in length
40007 40106: gap of 100 bp in length
40107 40967: contig of 861 bp in length
40968 41067: gap of 100 bp in length
41068 41913: contig of 846 bp in length
41914 42013: gap of 100 bp in length
42014 42824: contig of 811 bp in length
42825 42924: gap of 100 bp in length
42925 43776: contig of 852 bp in length
43777 43876: gap of 100 bp in length
43877 44752: contig of 876 bp in length
44753 44852: gap of 100 bp in length
44853 45724: contig of 872 bp in length
45725 45824: gap of 100 bp in length
45825 46643: contig of 819 bp in length
46644 46743: gap of 100 bp in length
46744 47599: contig of 856 bp in length
47600 47699: gap of 100 bp in length
47700 48551: contig of 852 bp in length
48552 48651: gap of 100 bp in length
48652 49485: contig of 834 bp in length
49486 49585: gap of 100 bp in length
49586 50440: contig of 855 bp in length
50441 50540: gap of 100 bp in length
50541 51404: contig of 864 bp in length
51405 51504: gap of 100 bp in length
51505 52372: contig of 868 bp in length
52373 52472: gap of 100 bp in length
52473 53328: contig of 856 bp in length
53329 53428: gap of 100 bp in length
53429 54268: contig of 840 bp in length
54269 54368: gap of 100 bp in length
54369 55229: contig of 861 bp in length
55230 55329: gap of 100 bp in length
55330 56197: contig of 868 bp in length
56198 56297: gap of 100 bp in length
56298 57163: contig of 866 bp in length
57164 57263: gap of 100 bp in length
57264 58130: contig of 867 bp in length
58131 59230: gap of 100 bp in length
59231 59082: contig of 852 bp in length
59083 59182: gap of 100 bp in length
59183 60020: contig of 838 bp in length
60021 60120: gap of 100 bp in length
60121 60983: contig of 863 bp in length
60984 61083: gap of 100 bp in length
61084 61935: contig of 852 bp in length
61936 62035: gap of 100 bp in length
62036 62866: contig of 831 bp in length
62867 62966: gap of 100 bp in length
62967 63827: contig of 861 bp in length
63828 63927: gap of 100 bp in length
63928 64783: contig of 856 bp in length
64784 64883: gap of 100 bp in length
64884 65740: contig of 857 bp in length
65741 65840: gap of 100 bp in length
65841 66684: contig of 844 bp in length
66685 66784: gap of 100 bp in length
66785 67651: contig of 867 bp in length
67652 67751: gap of 100 bp in length

```

Query Match 48.2%; Score 271.6; DB 2: length 65570;
 Best Local Similarity 89.8%; Pred. No. 9e-55;

```

Matches 289; Conservative C; Mismatches 33; Indels 0; Gaps 0;
QY 243 CCCACCCACCCACTCCCCCACTCCCTAGATCTGCTGGGGGCGGCGTGAACCTGGC 302
DB 67642 CCCCCNCCCCACCACTCCCCCAANTATAAGATGTGTCGFGGGCGGTGCACTGGC 67583
QY 303 CTTTTAAGGGGGGGGGGGGGCTCCACGTCGCTTTCTGCTGAGTGAAGTGAACATAAAC 362
DB 67582 CATAAAGGGGGGGGGGGGGCTACACGCTTTCGCTGAGTGAAGTGAACATAAAC 67523
QY 363 AGAGCCGGGAAAGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 422
DB 67522 AGAGCCGGGAAAGGGGGGGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 67463
QY 423 GCGCTCGTGCGAGCGCAATCTATAAAGGAAGTCTCCGGGCAAAAACCCGTAATTCG 482
DB 67462 GCGCTCGTGCGAGCGCAATCTATAAAGGAAGTCTCCGGGCAAAAACCTCTAATTCG 67403
QY 483 AGCGAGAGTGAAGTGGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 542
DB 67402 AGCGAGAGTGAAGTGGGGGGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 67343
QY 543 GCGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 564
DB 67342 GCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 67321

RESULT 14
AC021345/c
LOCUS Homo sapiens clone Rpl1-24J9, Low-PASS SEQUENCE SAMPLING.
DEFINITION AC021345
ACCESSION AC021345
VERSION AC021345.2 GI:9130845
KEYWORDS HTG: HTGS-PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 90698)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone Rpl1-24J9
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 90698)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beca,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Donico,M., Doyle,M., Fenestor,J.,
Ferrelia,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisanl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6705761.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RX/RepeatMasker.html
Smit, A.F.A. & Green, P. (1996-1997)
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project information

```

Center project name: L4483
Center clone name: 24_J_9

* NOTE: This record contains 92 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 910: contig of 910 bp in length
* 911 1020: gap of 100 bp
* 1011 1873: contig of 863 bp in length
* 1874 1973: gap of 100 bp
* 1974 2824: contig of 851 bp in length
* 2825 2924: gap of 100 bp
* 2925 3802: contig of 878 bp in length
* 3803 3902: gap of 100 bp
* 3903 4816: contig of 914 bp in length
* 4817 4916: gap of 100 bp
* 4917 5739: contig of 843 bp in length
* 5760 5859: gap of 100 bp
* 5860 6764: contig of 905 bp in length
* 6765 6864: gap of 100 bp
* 6865 7747: contig of 883 bp in length
* 7748 7847: gap of 100 bp
* 7848 8755: contig of 908 bp in length
* 8756 8855: gap of 100 bp
* 8856 9763: contig of 898 bp in length
* 9754 9853: gap of 100 bp
* 9854 10757: contig of 904 bp in length
* 10758 10857: gap of 100 bp
* 10858 11732: contig of 875 bp in length
* 11733 11832: gap of 100 bp
* 11833 12739: contig of 907 bp in length
* 12740 12839: gap of 100 bp
* 12840 13710: contig of 871 bp in length
* 13711 13810: gap of 100 bp
* 13811 14684: contig of 874 bp in length
* 14685 14784: gap of 100 bp
* 14785 15662: contig of 878 bp in length
* 15663 15762: gap of 100 bp
* 15763 16677: contig of 915 bp in length
* 16678 16777: gap of 100 bp
* 16778 17678: contig of 901 bp in length
* 17679 17778: gap of 100 bp
* 17779 18679: contig of 901 bp in length
* 18680 18779: gap of 100 bp
* 18780 19632: contig of 853 bp in length
* 19633 19732: gap of 100 bp
* 19733 20634: contig of 902 bp in length
* 20635 20734: gap of 100 bp
* 20735 21620: contig of 886 bp in length
* 21621 21720: gap of 100 bp
* 21721 22573: contig of 859 bp in length
* 22580 22679: gap of 100 bp
* 22680 23568: contig of 889 bp in length
* 23569 23668: gap of 100 bp
* 23669 24554: contig of 866 bp in length
* 24555 24654: gap of 100 bp
* 24655 25521: contig of 867 bp in length
* 25522 25621: gap of 100 bp
* 25622 26487: contig of 866 bp in length
* 26488 26581: gap of 100 bp
* 26588 27464: contig of 877 bp in length
* 27465 27564: gap of 100 bp
* 27565 28466: contig of 902 bp in length
* 28467 28566: gap of 100 bp
* 28567 29464: contig of 898 bp in length

29465 29564: gap of 100 bp
* 29565 30447: contig of 883 bp in length
* 30448 30547: gap of 100 bp
* 30548 31453: contig of 906 bp in length
* 31454 31553: gap of 100 bp
* 31554 32452: contig of 899 bp in length
* 32453 32552: gap of 100 bp
* 32553 33447: contig of 895 bp in length
* 33448 33547: gap of 100 bp
* 33548 34435: contig of 888 bp in length
* 34436 34535: gap of 100 bp
* 34536 35433: contig of 898 bp in length
* 35434 35534: gap of 100 bp
* 35534 36440: contig of 907 bp in length
* 36441 36540: gap of 100 bp
* 36541 37422: contig of 882 bp in length
* 37423 37522: gap of 100 bp
* 37523 38402: contig of 880 bp in length
* 38403 38502: gap of 100 bp
* 38503 39380: contig of 878 bp in length
* 39381 39480: gap of 100 bp
* 39481 40407: contig of 927 bp in length
* 40408 40507: gap of 100 bp
* 40508 41405: contig of 898 bp in length
* 41406 41505: gap of 100 bp
* 41506 42399: contig of 894 bp in length
* 42400 42499: gap of 100 bp
* 42500 43357: contig of 858 bp in length
* 43358 43457: gap of 100 bp
* 43458 44356: contig of 899 bp in length
* 44357 44456: gap of 100 bp
* 44457 45325: contig of 869 bp in length
* 45326 45425: gap of 100 bp
* 45426 46305: contig of 880 bp in length
* 46306 46405: gap of 100 bp
* 46406 47302: contig of 887 bp in length
* 47303 47402: gap of 100 bp
* 47403 48293: contig of 891 bp in length
* 48294 48393: gap of 100 bp
* 48394 49280: contig of 887 bp in length
* 49281 49380: gap of 100 bp
* 49381 50250: contig of 870 bp in length
* 50251 50350: gap of 100 bp
* 50351 51209: contig of 859 bp in length
* 51210 51309: gap of 100 bp
* 51310 52221: contig of 912 bp in length
* 52222 52321: gap of 100 bp
* 52322 53205: contig of 884 bp in length
* 53206 53305: gap of 100 bp
* 53306 54225: contig of 920 bp in length
* 54226 54325: gap of 100 bp
* 54326 55206: contig of 881 bp in length
* 55207 55306: gap of 100 bp
* 55307 56165: contig of 859 bp in length
* 56166 56265: gap of 100 bp
* 56266 57124: contig of 859 bp in length
* 57125 57224: gap of 100 bp
* 57225 58121: contig of 897 bp in length
* 58122 58221: gap of 100 bp
* 58222 59075: contig of 854 bp in length
* 59076 59175: gap of 100 bp
* 59176 60058: contig of 883 bp in length
* 60059 60158: gap of 100 bp
* 60159 61067: contig of 909 bp in length
* 61068 61167: gap of 100 bp
* 61168 62051: contig of 884 bp in length
* 62052 62151: gap of 100 bp
* 62152 63022: contig of 871 bp in length
* 63023 63122: gap of 100 bp
* 63123 64026: contig of 904 bp in length
* 64027 64126: gap of 100 bp
* 64127 65015: contig of 889 bp in length
* 65016 65115: gap of 100 bp

[illegible]

RESULT 15	AF258627	LOCUS	AF258627	697 bp	mRNA	linear	PR1 11-MAY-2000
		DEFINITION	Homo sapiens ATP binding cassette transporter 1 (ABCA1) mRNA, partial cds.				

AF258627
GI:7769707

Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 697)
Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C.,
Acouzerat,B.E., Fielding,C.J. and Kane,J.P.
Analysis of hABC1 gene 5' end: additional peptide sequence,
promoter region, and four polymorphisms
Biochem. Biophys. Res. Commun. 271 (2000) In press

2 (bases 1 to 697)
Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C.,
Acouzerat,B.E., Fielding,C.J. and Kane,J.P.
Direct Submission
Submitted (19-APR-2000) Cardiovascular Research Institute,
University of California, San Francisco, 505 Parnassus Avenue, San
Francisco, CA 94143-0130, USA

```

FEATURES
    source              Location/Qualifiers
1..697
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="9"
    /map="9q31"
    /tissue_type="placenta"
gene
1..>697
    /gene="ABCA1"
    /note="ABCA1"
396..>697
    /gene="ABCA1"
    /note="membrane-bound"
    /codon_start=1
    /product="ATP binding cassette transporter 1"
    /protein_id="AAF69513.1"
    /db_xref="GI:7769708"
    /translation="NACWQPLRLLLKWLNTFRRRQ*COLLEAVAMPFLIFLILISVRL
SPPPVEOHECHPNKAMPASAGTLPWQGIICNANNECFRYPPTCEAPGVGVGNENKS"
BASE COUNT      152 a   198 c   190 g   156 t   1 others
ORIGIN
Query Match      30.7%;   Score 173;   DB 9;   Length 697;
Best Local Similarity 100.0%;   Pred. No. 3.8e-31;
Matches 173;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;
QY 392 AGAGCACAGCGTTTGACCGGATAGTAACCTCTGGCGTCGGTGCACCGCAATCTATAAAGG 451
    |||||
Db 1 AGAGCACAGCGTTTGACCGGATAGTAACCTCTGGCGTCGGTGCACCGCAATCTATAAAGG 60
QY 452 ACTAGTCCCGGAAAAACCCGTAATTCGACGAGAGTGAGTGGGCGGGGACCCGCA 511
    |||||
Db 61 AACTAGTCCCGGAAAAACCCGTAATTCGACGAGAGTGAGTGGGCGGGGACCCGCA 120
    |||||
QY 512 GAGCGGAGCGGACCCCTCTCTCCCGGGCTCGGCGAGGCGCAGGGCGGGAGGTC 564
    |||||
Db 121 GAGCGGAGCGGACCCCTCTCTCCCGGGCTCGGCGAGGCGCAGGGCGGGAGGTC 173
    |||||

Search completed: February 4, 2003, 01:21:04
Job time : 2000.75 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 16:39:45 ; Search time 1067.18 Seconds

(without alignments)
8559.210 Million cell updates/sec

Title: US-09-596-141c-3_COPY_1080_1643

Perfect score: 564

Sequence: 1 gagctctctctctccccaac.....cagggcaggcgaggagctc 564

Scoring table: IDENTITY_KUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estm:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_estc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hic:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_nam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89.4	15.9	736	9	AU135588
2	89	15.8	547	9	AL698654
3	89	15.8	763	9	AU121731
4	75	13.3	292	14	244377
5	75	13.3	1004	11	BC034824
6	60.6	10.8	1261	17	CNS015W3

7	60.2	10.7	998	12	BG678861
8	54.8	9.7	298	17	CNS04M2X
9	53.8	9.5	844	17	CNS0052P
c 10	52.4	9.3	910	17	CNS0060N
c 11	51.2	9.1	832	17	CNS015XC
c 12	50.6	9.0	925	17	CNS0091P
c 13	50.6	9.0	1101	17	CNS017SY
c 14	50.4	8.9	870	17	CNS006E2
c 15	49.8	8.8	844	17	CNS0052P
c 16	48.4	8.6	681	17	CNS02FE9
c 17	48.4	8.6	1004	14	BQ944204
c 18	48	8.5	900	17	AG081217
c 19	47.6	8.4	1609	17	CNS010EW
c 20	47.6	8.4	1101	17	CNS01720
c 21	47.4	8.4	1059	14	BQ680479
c 22	47.4	8.4	1101	17	CNS017V2
c 23	47.2	8.4	477	9	AL513813
c 24	47.2	8.4	925	17	CNS0091P
c 25	47.2	8.4	942	17	AG042878
c 26	47	8.3	843	17	CNS00CS1
27	46.8	8.3	431	9	AL513947
28	46.2	8.2	527	9	AL514325
29	46.2	8.2	875	17	AG043475
c 30	46.2	8.2	1136	14	BQ943816
c 31	46.2	8.2	1417	14	BM810236
c 32	46	8.2	758	17	AG073529
c 33	45.8	8.1	523	9	AL514015
c 34	45.8	8.1	766	17	AG041031
c 35	45.8	8.1	932	17	CNS0072Q
c 36	45.8	8.1	997	14	BQ953024
37	45.6	8.1	456	9	AL554951
38	45.6	8.1	803	17	AG157500
c 39	45.6	8.1	1213	13	BM560246
40	45.4	8.0	426	17	CNS00A3H
41	45.2	8.0	853	14	BQ219067
c 42	45	8.0	683	12	BG086652
c 43	45	8.0	884	17	CNS018NP
c 44	45	8.0	1129	14	BM914614
45	45	8.0	1151	17	CNS024TJ

ALIGNMENTS

RESULT 1	AU135588	736 bp	mRNA	linear	EST 02-AUG-2002
LOCUS	AU135588	PLACEL	Homo sapiens	cdna	clone PLACE1002437.5, mRNA
DEFINITION	AU135588	sequence.			
ACCESSION	AU135588	GI:10996127			
VERSION	AU135588.1				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 736)				
AUTHORS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
	Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y., Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and Isozaki, T.				
TITLE	HRI human cDNA project				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Takao Isozaki				
	Genomics Laboratory				
	Helix Research Institute				
	1532-3 Yana, Kisarazu, Chiba 292-0812, Japan				
	Tel: 81-438-52-3975				
	Fax: 81-438-52-3986				
	Email: genomics@hri.co.jp				
	HRI human cDNA project: 5'- & 3'-end one pass sequencing; Helix Research Institute; cDNA library construction; Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.				

AUTHORS
 Aubiffay,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
 M.D., Duprat,S., Houlgatte,R., Jumeau,M.M., Lamy,B., Lorenzo,F.,
 Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
 Sebastiani-Kabatchis,C. and Tessier,A.

TIIE
 IMAGE: molecular integration of the analysis of the human genome
 and its expression

JOURNAL
 C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

MEDLINE
 95277534

COMMENT
 Contact: Genethon
 Genethon
 Genethon Centre de recherche sur le Genome Humain
 1 rue de l'Internationale, BP60 92002 EVRY Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr

Single read
 Genexpress_library_id: C; Genexpress_sequence_id: ylc-lzb08
 Seq primer: (-21)M13_universal.

FEATURES
 source
 ..292
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="c-lzb08"
 /clone_lib="normalized infant brain cDNA"
 /sex="Female"
 /tissue_type="total brain"
 /dev_stage="3 months old"
 /note="Organ: brain; Vector: lafmid BA; Site: HindIII;
 Site2: NotI; sex:female; dev_stage:3 months old;
 isolate: muscular atrophy patient; tissue_type: total brain
 ; total mRNA was oligo-(dT) primed and directionally
 cloned 5' -> 3' into the HindIII -> NotI sites of the
 lafmid BA vector. Clone library from B. Soares, Psychiatry
 Dept., Columbia University, USA. Normalization method:
 Bento Soares, P.N.A.S in press"

BASE COUNT
 50 a 87 c 96 g 56 t 3 others

ORIGIN
 Query Match 13.3%; Score 75; DB 14; Length 292;
 Best Local Similarity 100.0%; Pred. No. 4.4e-07;
 Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 490 GTGAGTGGGGCGGAGCCGAGAGCCGAGCCGAGCCCTCTCTCCCGGCTCGGCGACGG 549
 |||||
Db 1 GTGAGTGGGGCGGAGCCGAGAGCCGAGCCGAGCCCTCTCTCCCGGCTCGGCGACGG 60

QY 550 CAGGCGGGGAGCTC 564
 |||||
Db 61 CAGGCGGGGAGCTC 75

RESULT 5
 BC034824 1004 bp mRNA linear HTC 26-JUL-2002

LOCUS
 BC034824 Homo sapiens, clone IMAGE:4749735, mRNA.

DEFINITION
 BC034824

ACCESSION
 BC034824.1 GI:21961568

VERSION
 HTC.

KEYWORDS
 Homo sapiens.

SOURCE
 Homo sapiens

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1004)

REFERENCE
 Strausberg,R.

AUTHORS
 Direct Submission

JOURNAL
 Submitted (24-JUL-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK
 NIH-MGC Project URL: http://mgc.nci.nih.gov

COMMENT
 Contact: MGC help desk
 Email: qcqaps-remail.nih.gov
 Tissue Procurement: James Cleaver, M.D.

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Haie, S.M.,
 Yoon, V.S., Kovis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Series: IRAX Plate: 52 Row: d Column: 19
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA y1: 5915657
 This clone has the following problem: incomplete processing.

FEATURES
 Location/Qualifiers
 ..1004
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4749735"
 /tissue_type="Skin, squamous cell carcinoma"
 /clone_lib="NCI_CGAP_Skn4"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6-ccdb"

BASE COUNT
 306 a 234 c 230 g 234 t

ORIGIN
 Query Match 13.3%; Score 75; DB 11; Length 1004;
 Best Local Similarity 100.0%; Pred. No. 3.4e-07;
 Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 490 GTGAGTGGGGCGGAGCCGAGAGCCGAGCCGAGCCCTCTCTCCCGGCTCGGCGACGG 549
 |||||
Db 1 GTGAGTGGGGCGGAGCCGAGAGCCGAGCCGAGCCCTCTCTCCCGGCTCGGCGACGG 60

QY 550 CAGGCGGGGAGCTC 564
 |||||
Db 61 CAGGCGGGGAGCTC 75

RESULT 6
 CANS015W3 1201 bp DNA linear GSS 26-JUL-1999

LOCUS
 CANS015W3 Drosophila melanogaster genome survey sequence S96 end of BAC
 BACN15123 of DrosBAC library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.

ACCESSION
 AL105381

VERSION
 AL105381.1 GI:5619397

KEYWORDS
 GSS.

SOURCE
 Drosophila melanogaster.

ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1201)

REFERENCE
 Genoscope.

AUTHORS
 Direct Submission

TITLE
 Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

JOURNAL
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the European Drosophila Genome Project (EDGP) -
 http://www.edgp.ebi.ac.uk - . This Drosophila melanogaster BAC
 library (Dros BAC) was made by Alain Billaut at CEPH (Centre
 d'Etude du Polymorphisme Humain) with funding provided by a MRC
 project grant. The DNA was prepared from embryos by Alain Bucheton
 and Genevieve Payan. It has been constructed in the vector
 pBelosAC11.

FEATURES
 Location/Qualifiers
 ..1201
 /organism="Drosophila melanogaster"

```
/db_xref="taxon:7227"
/clone="BACN15123"
/clone_lib="DrosBAC"
/plasmid="pBeLoBAC11"
/note="end : SP6"

BASE COUNT      255 a      276 c      278 g      235 t      157 others
ORIGIN

Query Match
Best Local Similarity 10.8%; Score 60.8; DB 17; Length 1201;
Matches 109; Conservative 51; Mismatches 132; Indels 0; Gaps 0;

QY 36 GGAAGTAAACAAAGAAAGAAATTCGGAAAGCAGGATTATAGAGAGCAAAATTCAC 95
Db 857 KRAAAMAAKKTATMTTGKATMKDTCGGMAATAAGAAAGAAWADDADTTKKK 916
QY 96 TGGTCCCTGCTCCGGGAACATGGACTAGAGATCTGGCGCAGCCCGAGCCAG 155
Db 917 TKTGGKSGSGSCCTSCCGSSCARASMGSGSGSGGCGCCMGCGCCGSGSCNGC 976
QY 156 CGCTCCCGCGCTTTAGGCGGGGCGCGCGGGGGAAGGGGACGCGACCGCGGA 215
Db 977 CGSSCCGCCNCGSSSSSGGGGGGGGCGGGGGGGGGGGGGGGGGGGGGGGGG 1036
QY 216 CCTAAGACACCTGTGTACCTTCACCCGCCACCCACCCACCTCCCCCACTCCCTAG 275
Db 1037 CCCSCCCCCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1096
QY 276 ATGTCTGTGGCGCTCAAGCTCCCGCTTAAGGGGGGGGGGGGGGGGGGGGGCTCC 327
Db 1097 CCSSSSSCSGGGGGGGSGCGSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1148

RESULT 7
LOCUS BG578861
DEFINITION 602624780f1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4749735 5',
mRNA sequence.
ACCESSION BG578861
VERSION BG578861.1 GI:13910258
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 998)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rstraus@nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10603 row: 9 column: 16
High quality sequence stop: 860.
Location/Qualifiers
1..998
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4749735"
/clone_lib="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: skin; Vector: pCMV-Sport6; Site: 1; Not1;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT      285 a      233 c      244 g      236 t

Query Match
Best Local Similarity 9.7%; Score 54.8; DB 17; Length 298;
Matches 81; Conservative 33; Mismatches 66; Indels 0; Gaps 0;

QY 101 CCCTTGGTCCGGGACGTGGACTAGAGTCTCGCGGAGCCCGGAGCCAGCCAGCTT 160
Db 82 CCNKASCCCVGVGCCVCSGCCSGSGSGSCSCSCSSSSSCSCSCSCSCSCSCSCSCSC 141
QY 161 CCCGCGGTCTTAGCCGGCGGGCGGGCGGGGAAGGGGACGACCGGGGACCTA 220
Db 142 GSGSSCCSSGSGSGSGGCGCCGCCSSSVSGSGSGSGGCGCCGCCGCGCGCNATC 201

ORIGIN

Query Match
Best Local Similarity 10.7%; Score 60.2; DB 12; Length 998;
Matches 65; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 492 GAGTGGGGCGGGACCGGAGCGGAGCGACCTTCTCTCCCGGGCTGGGGCAGGGCA 551
Db 1 GTGAGTGGGGGGGACCGGAGCGGAGCGGAGCGACCTTCTCTCCCGGGCTGGGGCAGGGCA 60
QY 552 GGGCGGGGAGCTC 564
Db 61 GGGCGGGGAGCTC 73

RESULT 8
LOCUS CNS04M2X
DEFINITION 298 bp DNA linear GSS 21-MAY-2000
Tetraodon nigroviridis genome survey sequence T7 end of clone
119N17 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL296898
VERSION AL296898.1 GI:8035478
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 298)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissensbach, J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 298)
AUTHORS Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissensbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 298)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
Location/Qualifiers
1..298
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="119N17"
/clone_lib="G"
/note="Genoscope sequence ID : C08G119CG09LP1-end : T7"

BASE COUNT      18 a      106 c      80 g      15 t      79 others
ORIGIN

Query Match
Best Local Similarity 40.5%; Score 54.8; DB 17; Length 298;
Matches 81; Conservative 33; Mismatches 66; Indels 0; Gaps 0;

QY 101 CCCTTGGTCCGGGACGTGGACTAGAGTCTCGCGGAGCCCGGAGCCAGCCAGCTT 160
Db 82 CCNKASCCCVGVGCCVCSGCCSGSGSGSCSCSCSSSSSCSCSCSCSCSCSCSCSCSC 141
QY 161 CCCGCGGTCTTAGCCGGCGGGCGGGCGGGGAAGGGGACGACCGGGGACCTA 220
Db 142 GSGSSCCSSGSGSGSGGCGCCGCCSSSVSGSGSGSGGCGCCGCCGCGCGCNATC 201
```

```

QY 221 AGACACCTGCTGTACCTCCACCCACCCACCCACCCACCTCCCGCAACTCCCTAGACTG 280
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 202 NNCNCNCCTCTCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNCNC 261
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 281 TGTGGGGGGCTGAACGTCG 300
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 252 GGGGGGGGGGGGGGGGGGGG 281

RESULT 9
CNS0052P
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence TET3 end of BAC #
  BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL065652
VERSION
  AL065652.1 GI:4932342
KEYWORDS
  GSS.
SOURCE
  Drosophila melanogaster.
  Drosophila melanogaster.
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 844)
REFERENCE
  Direct Submission
  AUTHORS
  TITLE
  JOURNAL
COMMENT
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the Berkeley Drosophila Genome Project (BDGP).
  The BDGP is constructing a physical map of the Drosophila
  melanogaster genome using these BACs. For further information
  please see http://www.fruitfly.org The BDGP Drosophila
  melanogaster BAC library was prepared by Kazutoyo Osoegawa and
  Aaron Mammoler in Pieter de Jong's laboratory in the Department of
  Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
  NY. The library is named RPCI-98 and was constructed by partial
  EcoRI digestion of Drosophila DNA provided by the BDGP from the
  isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
  p1 and EST libraries. A more detailed description of the library
  and how to order individual BAC clones, the entire library, or
  filters for hybridization from the BACPAC Resource Center can be
  found at http://bacpac.med.buffalo.edu/drosophila_bac.h.m.
FEATURES
  source
  1..844
  Location/Qualifiers
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone="BACR14J21"
    /clone_lib="RPCI-98"
    /note="end : TET3"
BASE COUNT 261 a 112 c 92 g 35 t 344 others
ORIGIN
  Query Match 9.5%; Score 53.8; DB 17; Length 844;
  Best Local Similarity 23.8%; Pred. No. 0.023;
  Matches 86; Conservative 131; Mismatches 140; Indels 5; Gaps 1;
QY 38 AAATAACAAGGAAAAAAATTCGGGAAGCAGGATTAGAGCAATTCCTACTG 97
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 AAAAAAAGAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 360
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 98 GTGCCCTTGGCTCGCGGAAGCTGACTAGAGTCTCGCGCGCAGCGCCGAGCCAGCG 157
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 361 SASVSSACSSVAGSMAAGACGCGCAANNACGCGSSSSASFSRMMSSCSSCASSSCS 420
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 158 CTTCCCGCGCTCTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 217
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 421 VGSSSSAMSSSCCVYSSCGMASSSCGMASSSCGMASSSCGMASSSCGMASSSCGMAS 480
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 218 CTAAGACACCTGCTGTACCTCCACCCACCCACCCACCCACCTCCCGCCAGCTCTAGAT 277
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 481 S-----MVCCCMCMCCMCCSVSVCAVCSGSGSVSVAVGCSVGGRRVGGSSRRAGRS 535
QY 278 GTTCGTGGCGGCTGAAGCTCCCGCTTTAAAGGGCGGGCGCCGCTCCACGTCTGCTTC 337
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 536 SCRGSGSSVSSGWSSSSVGMCACSSASVSCBSVSVSSVSGBVRGRCRCVGGGV 595
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 338 TGCTGAGTGAAGTAACTACATAACAGAGCGCGGAGCGGCGGAGGAGGAGGAGCA 397
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 596 GGSRVSSCGSSSSSSSSSSCGSVSRRSRSGANGVYVGGSGRRGGGGGRRGGANARAD 655
QY 398 CA 399
Db 656 DA 657

RESULT 10
CNS0060N/c
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence 17 end of BAC #
  BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL065629
VERSION
  AL065629.1 GI:4944698
KEYWORDS
  GSS.
SOURCE
  Drosophila melanogaster.
  Drosophila melanogaster.
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 910)
REFERENCE
  Direct Submission
  AUTHORS
  TITLE
  JOURNAL
COMMENT
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the Berkeley Drosophila Genome Project (BDGP).
  The BDGP is constructing a physical map of the Drosophila
  melanogaster genome using these BACs. For further information
  please see http://www.fruitfly.org The BDGP Drosophila
  melanogaster BAC library was prepared by Kazutoyo Osoegawa and
  Aaron Mammoler in Pieter de Jong's laboratory in the Department of
  Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
  NY. The library is named RPCI-98 and was constructed by partial
  EcoRI digestion of Drosophila DNA provided by the BDGP from the
  isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
  p1 and EST libraries. A more detailed description of the library
  and how to order individual BAC clones, the entire library, or
  filters for hybridization from the BACPAC Resource Center can be
  found at http://bacpac.med.buffalo.edu/drosophila_bac.h.m.
FEATURES
  source
  1..910
  Location/Qualifiers
    /organism="Drosophila melanogaster"
    /db_xref="taxon:7227"
    /clone="BACR14J21"
    /clone_lib="RPCI-98"
    /note="end : 17"
BASE COUNT 202 a 63 c 112 g 198 t 335 others
ORIGIN
  Query Match 9.3%; Score 52.4; DB 17; Length 910;
  Best Local Similarity 20.5%; Pred. No. 0.047;
  Matches 48; Conservative 104; Mismatches 62; Indels 0; Gaps 0;
QY 39 AACTAACAAGGAAAAAAATTCGGGAAGCAGGATTAGAGCAATTCCTACTG 98
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 752 AASARVAAVVAVVAASAAVVAASAAVVAASAAVVAASAAVVAASAAVVAASAAVVA 693
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 99 TGCCTTGGCTCGCGGAAGCTGACTAGAGTCTCGCGCGCAGCGCCGAGCCAGCCG 158
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 692 SASSSASASMSVSCRSAVSAASAAASASASASASASASASASASASASASAS 633
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 159 TTCCCGCGCTCTTAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 218

```



```

RESULT 13
CNS017SY
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108460
VERSION AL108460.1 GI:5628764
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (BDGP) -
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
Location/Qualifiers
source
BASE COUNT 254 a 176 c 160 g 152 t 359 others
ORIGIN
Query Match 9.0%; Score 50.6; DB 17; Length 1101;
Best Local Similarity 13.4%; Pred. No. 0.12;
Matches 53; Conservative 187; Mismatches 157; Indels 0; Gaps 0;

Oy 129 GAGTCTGGGGGACCGGACCGGCTTCGGGGGCTTACGGCGGCGGCGGCGG 188
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 675 GKKAKGVRGVRCCAGGGGACACMAADCGCCAKMACCGSSSSASSTSSSSAS 734
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Oy 189 GCGGGGAGGAGGACGACCGGACCCCTAAGACACCTCTGTACCTCCACCCGAC 248
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 735 RGVVSCACSGSGGASAGSAGGGGGRGGGCGCCASGGVCGAACSSASCSMSMC 794
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Oy 249 CCCACCCACCTCCCGCAACCTCCCTAGATGTGTGTCGGCGGCTGAAGCTGCCGCTTA 308
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 795 GVSGGSCSASGCGCGVSSCSAYASASSASVMSKVASAVASCSAVASGAGVSSCR 854
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Oy 309 AGGGGGGGGCGGCGGCTTCTCTCTCTGAGTACTGAATAAAGAGAGG 368
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 855 SSVASSVSAASVSSVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 914
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Oy 369 CGGAACGGGGGAGGAGGAGACAGAGCTTTCACCGATAGTAACTCTCCGCTC 428
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 915 VVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 974
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Oy 429 GGTGACCGCAATCTATAAAGGAAGTACTGTCGGGCAAAAACCGGTAAATTCGCGGAG 488
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 975 ASSSSSVSVSVVAVASAAAAAASSSSSASAVAVSVASSSSSSSSSSSSSSS 1034
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Oy 489 AGTGAGTGGGCGGCGGACCGACGCGGACCGGACCG 525
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 1035 SVSSSSSSSSSVSVSSSSSVAVASASASVSSV 1071
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
RESULT 14

```

```

CNS006EZ/c
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR13E15 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL064271
VERSION AL064271.1 GI:4944346
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 870)
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osadaawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
source
BASE COUNT 256 a 90 c 87 g 127 t 310 others
ORIGIN
Query Match 8.9%; Score 50.4; DB 17; Length 870;
Best Local Similarity 28.0%; Pred. No. 0.13;
Matches 67; Conservative 74; Mismatches 98; Indels 0; Gaps 0;

Oy 95 CIGTGCCCTTGGTGC CGGGAAGGAGTACTAGAGTCTCGGCGAGCCCGAGCCCA 154
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 809 CSSGTGTTCTSSCTTTGCGCKSSGGTSSSCTBTTCGSCSTCSSSTCTCTTCG 750
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Oy 155 CGCTTCCCGCGCTCTTAGCCCGGCGGCGGCGGGAAGGACGACGACCGCGG 214
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 749 GCSSTYSSSTTTCBKSSSGSXAANKVSGSGGCSYVVGASMSGCGRASAGKRS 690
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Oy 215 ACCCTAACACACCTGCTGTACCTCCACCCACCCACCTCCCTCCCTCCCTA 274
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 689 ASSGHGCSGCGSGSGSYSACCCMCSCSCYCCSCSYCCSCCTCKTTTGBBSC 630
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Oy 275 GATGTGTCGTGGCGGCGTGAAGCTGCGCCGTTTAAGGGGCGGCGGCGGCTCCAGTGC 333
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 629 GCTSBKSMGMSGGNGTCYCCYCCCTCTCCGSCYSCGCGGCGGCTCCSCGSCSY 571
| : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :

RESULT 15
CNS0052P
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR1P16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL056652
VERSION AL056652.1 GI:4932342

```

Search completed: February 4, 2003, 03:26:21
Job time : 1080.9 secs

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 12:56:40 ; Search time 26.4714 Seconds
(without alignments)
5534.066 Million cell updates/sec

Title: US-09-596-141C-3_COPY_1080_1643

Perfect score: 564

Sequence: 1 gagctctctctcccccaatc.....cagggcaggcgaggagcgc 564

Scoring table: IDENTITY_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first: 45 summaries

Database :

Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38.8	6.9	595	4	US-08-483-533-4
2	38.8	6.9	595	4	US-09-283-471A-4
3	38.8	6.9	1327	4	US-08-483-533-36
4	38.8	6.9	1327	4	US-08-283-471A-36
5	38.8	6.9	1335	5	PCT-US91-06532-1
6	38.6	6.8	4481	4	US-09-041-886-18
7	37.6	6.7	936	1	US-08-018-977C-4
8	36.2	6.4	43280	2	US-08-804-237C-1
9	35.8	6.3	4085	1	US-08-072-574-7
10	35.8	6.3	4181	1	US-08-072-574-9
11	35.8	6.3	11907	4	US-08-061-376-4
12	35.6	6.3	4411529	4	US-09-103-840A-1
13	35.6	6.3	4411529	4	US-09-103-840A-1
14	35.2	6.2	35100	1	US-08-306-691B-19
15	35.2	6.2	35100	5	PCT-US93-06251-19
16	35	6.2	7218	1	US-08-232-463-14
17	34.6	6.1	1280	4	US-08-483-533-38
18	34.6	6.1	1280	4	US-09-283-471A-38
19	34.6	6.1	17410	1	US-07-841-646-3
20	34.6	6.1	17410	1	US-08-147-023-3
21	34.6	6.1	17410	1	US-08-477-570-3
22	34.6	6.1	17410	2	US-08-449-700-3
23	34.6	6.1	17410	2	US-08-449-699A-3
24	34.4	6.1	954	1	US-08-170-294-1
25	34.4	6.1	954	2	US-08-664-855-1
26	34.4	6.1	954	3	US-09-049-289-1
27	34.4	6.1	2888	4	US-08-765-907A-1

C 28	34.2	6.1	1490	4	US-08-964-127-3	Sequence 3, Appli
C 29	34.2	6.1	1490	4	US-09-496-692-3	Sequence 3, Appli
C 30	34.2	6.1	1610	4	US-09-276-531-56	Sequence 56, Appli
C 31	34.2	6.1	2460	4	US-08-964-127-1	Sequence 1, Appli
C 32	34.2	6.1	2460	4	US-09-496-692-1	Sequence 1, Appli
C 33	34.2	6.1	4085	1	US-08-486-270-7	Sequence 7, Appli
C 34	34.2	6.1	4085	3	US-08-367-264-7	Sequence 7, Appli
C 35	34.2	6.1	4085	4	US-09-153-757-7	Sequence 7, Appli
C 36	34.2	6.1	4181	1	US-08-486-270-9	Sequence 9, Appli
C 37	34.2	6.1	4181	3	US-08-367-264-9	Sequence 9, Appli
C 38	34.2	6.1	4181	4	US-09-153-757-9	Sequence 9, Appli
C 39	34.2	6.1	4207	4	US-08-660-148-1	Sequence 1, Appli
C 40	34.2	6.1	4207	4	US-08-660-148-3	Sequence 3, Appli
C 41	34.2	6.1	4303	4	US-08-660-148-4	Sequence 4, Appli
C 42	34.2	6.1	4303	4	US-08-660-148-6	Sequence 6, Appli
C 43	34.2	6.1	72604	4	US-09-268-932-7	Sequence 7, Appli
C 44	34.2	6.1	72604	4	US-09-657-474-7	Sequence 7, Appli
C 45	33.8	6.0	2877	4	US-09-235-103-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-483-533-4
; Sequence 4, Application US/08483533
; Patent No. 6172047
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; INVENTOR: Chou, Joany
; TITLE OF INVENTION: Method for Treating Tumorigenic
; DISEASES
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gersteitz, Murray & Brown
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentia Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,533
; FILING DATE: 07-MAR-95
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/419,853
; FILING DATE: 11-APR-95
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/861,233
; FILING DATE: 31-MAR-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, James P.
; REGISTRATION NUMBER: 28,491
; REFERENCE/DOCKET NUMBER: 28097/32742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-483-533-4

Query Match 6.9%; Score 38.8; DB 4; Length 595;

maps	0;
217	
8412	
277	

containing
phosphonates

```
; REFERENCE/DOCKET NUMBER: 9586
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS JR
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US93-06251-19

Query Match      6.28; Score 35.2; DB 5; Length 35100;
Best Local Similarity 50.0%; Pred. No. 4;
Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

Qy 158 CTTCCCGCGCGCTTTAGCGCGCGCGCGCGGGAAGGGGACGACGACCGCGGACC 217
    |||||
Db 8471 CCTGCCCCGACACCTTGTCTAGTCTCTGCCCATAGATGATCTGTGGACCATTTGAGGGGAAG 8412
    |||||

Qy 218 CTAAAGACACCTGTGTACCCCTCCACGCCACCCACCGACCTCCCCCAACTCCCTAGAT 277
    |||||
Db 8411 AAGTGAAAGAAAGAGAGACATCCCACTGCCCGCCCATCACACCCAGTCCCGCAGTCCACCC 8352
    |||||

Qy 278 GTGTGTTGGGGGCTGACGTCGCCCGTTTAAAGGGGGGGCGCGGCTCCACGTGC 333
    |||||
Db 8351 ATGTGGCCCGCGGCTCTCTGTGTCCTTTCAGCCCTGGCTGCGGGTCCCATGC 8296
    |||||
```

Search completed: February 3, 2003, 18:51:27
Job time : 8913.47 secs

GenCore version: 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 16:38:05 ; Search time 1325.59 Seconds
(without alignments)
10165.006 Million cell updates/sec

Title: US-09-596-141c-3_COPY_1181_1643

Perfect score: 463

Sequence: 1 ccttgctgcgcgggaactgtg.....caggggcaggggcgggagctc 463

Scoring table: IDENTITY_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmul: *

1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_cv: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_cv: *
23: em_pat: *
24: em_ph: *
25: em_pi: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pin: *
35: em_htg_rod: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	463	100.0	1167	9	HSA252201	AJ252201 Homo sapi
2	463	100.0	1643	6	AX060715	AX060715 Sequence
3	463	100.0	1643	6	AX060894	AX060894 Sequence
C 4	463	100.0	96717	9	AL359182	AL359182 Human DNA
5	463	100.0	175064	2	AC012230	AC012230 Homo sapi
6	459.8	99.3	1167	9	AF258623S1	AF258623 Homo sapi
7	448.8	96.9	69570	2	AC021246	AC021246 Homo sapi
8	447	96.5	183999	6	AX092589	AX092589 Sequence
9	446.4	96.4	3231	3	AX351029	AX351029 Sequence
10	446.4	96.4	149034	3	AF275948	AF275948 Homo sapi
-1	441.6	95.4	201144	5	AF287262	AF287262 Homo sapi
12	355.4	76.8	2893	6	AX351031	AX351031 Sequence
C 13	271.6	58.7	69570	2	AC021246	AC021246 Homo sapi
C 14	229.8	49.6	90698	2	AC021345	AC021345 Homo sapi
15	173	37.4	857	9	AF258627	AF258627 Homo sapi
C 16	155.4	33.6	186889	2	AL807243	AL807243 Mus muscu
17	155.4	33.6	278572	10	AF287263	AF287263 Mus muscu
C 18	153.2	33.1	145833	2	AC125837	AC125837 Rattus no
19	98	21.2	7260	6	AX253452	AX253452 Sequence
20	91	19.7	221	6	AX351032	AX351032 Sequence
21	89.4	19.3	1556	9	AK024328	AK024328 Homo sapi
22	89	19.2	1750	9	AK022254	AK022254 Homo sapi
23	75	16.2	9854	6	AX127831	AX127831 Sequence
24	75	16.2	9854	6	AX139818	AX139818 Sequence
25	67	14.5	10442	6	AX060713	AX060713 Sequence
26	67	14.5	10442	6	AX060892	AX060892 Sequence
27	67	14.5	10442	9	AF285167	AF285167 Homo sapi
28	60	13.0	10474	6	AX060719	AX060719 Sequence
29	60	13.0	10474	6	AX060721	AX060721 Sequence
30	60	13.0	10474	6	AX060898	AX060898 Sequence
31	60	13.0	10474	6	AX060900	AX060900 Sequence
C 32	49.6	10.7	96136	2	AC097674	AC097674 Rattus no
C 33	49.4	10.7	173127	2	AC131142	AC131142 Rattus no
C 34	49	10.6	65780	2	AC111741	AC111741 Rattus no
35	49	10.6	125020	9	AF429315	AF429315 Homo sapi
C 36	47.4	10.2	180461	2	AC105563	AC105563 Rattus no
C 37	47	10.2	167886	2	AC126316	AC126316 Rattus no
C 38	46.6	10.1	55357	2	AC121553	AC121553 Mus muscu
C 39	46.4	10.0	184402	2	AC127041	AC127041 Rattus no
C 40	46.2	10.0	150305	2	AC114705	AC114705 Rattus no
C 41	46.2	10.0	209216	2	AC117126	AC117126 Rattus no
42	46	9.9	134021	2	AC109884	AC109884 Rattus no
C 43	46	9.9	174096	2	AC097177	AC097177 Rattus no
C 44	45.6	9.8	110000	2	AC096324_0	AC096324 Rattus no
C 45	45.6	9.8	110000	2	AC096324_1	Continuation (2 of

ALIGNMENTS

RESULT 1
HSA252201
LOCUS
DEFINITION
Homo sapiens partial ABC-1 gene for ATP-binding cassette transporter-1, 5'UTR and promoter region.
ACCESSION
AJ252201
KEYWORDS
ABC-1 gene; ATP-binding cassette transporter-1; promoter.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1167)
AUTHORS
Porsch-Ozcurumez, M., Langmann, T., Heimerl, S., Borsukova, H.,
Kaminski, W.E., Drobniak, W., Honer, C., Schumacher, C. and Schmitz, G.


```

Db 17329 CTTTGACCATAGTAACTCTCGCTCGGTCGGCGACCCGAGTATATAAAGGAACIAGTCCC 17270
QY 361 GGCAGAAACCCCGTAAATTCGCGAGCGAGAGTGTAGTGGCGCGGAGCCGAGCGAGCC 420
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 17269 GGCAGAAACCCCGTAAATTCGCGAGCGAGAGTGTAGTGGCGCGGAGCCGAGCGAGCC 17210
QY 421 GACCCCTCTCTCCCGGCGTGGCGCAGGCGAGCGCGGAGCTC 463
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 17209 GACCCCTCTCTCCCGGCGTGGCGCAGGCGAGCGCGGAGCTC 17167

RESULT 5
AC012230
LOCUS
DEFINITION
Homo sapiens clone RP11-1M10, WORKING DRAFT SEQUENCE, 39 unordered
pieces.
AC012230
VERSION
AC012230.3 GI:7637254
KEYWORDS
HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 175064)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-1M10
Unpublished
2 (bases 1 to 175064)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgaiter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gace,D., Herten,L.,
Gallagan,J., Gardyna,S., Grant,S., Hages,S., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karates,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,X., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 22, 2000 this sequence version replaced gi:6454033.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RN/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2510
Center clone name: 1.M.10
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 117571 bases at least Q40
Consensus quality: 145749 bases at least Q30
Consensus quality: 160940 bases at least Q20
Insert size: 185000; agarose-fp
Insert size: 171264; sum-of-contigs
Quality coverage: 2.9 in Q20 bases; agarose-fp
Quality coverage: 3.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as

```

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1003: contig of 1003 bp in length
 1004 1103: gap of 100 bp
 1104 2634: contig of 1531 bp in length
 2635 2734: gap of 100 bp
 2735 4415: contig of 1681 bp in length
 4416 4515: gap of 100 bp
 4516 5785: contig of 1270 bp in length
 5786 5885: gap of 100 bp
 5886 7879: contig of 1994 bp in length
 7880 7979: gap of 100 bp
 7980 9686: contig of 1707 bp in length
 9687 9786: gap of 100 bp
 9787 12253: contig of 2467 bp in length
 12254 12353: gap of 100 bp
 12354 15228: contig of 2875 bp in length
 15229 15328: gap of 100 bp
 15329 17200: contig of 1872 bp in length
 17201 17300: gap of 100 bp
 17301 20131: contig of 2831 bp in length
 20132 20231: gap of 100 bp
 20232 22587: contig of 2356 bp in length
 22588 22687: gap of 100 bp
 22688 25707: contig of 3020 bp in length
 25708 25807: gap of 100 bp
 25808 28184: contig of 2377 bp in length
 28185 28284: gap of 100 bp
 28285 31338: contig of 3054 bp in length
 31339 31438: gap of 100 bp
 31439 34259: contig of 2861 bp in length
 34300 34399: gap of 100 bp
 34400 38318: contig of 3919 bp in length
 38319 38418: gap of 100 bp
 38419 42835: contig of 4417 bp in length
 42836 42935: gap of 100 bp
 42936 45448: contig of 2513 bp in length
 45449 45548: gap of 100 bp
 45549 48116: contig of 2568 bp in length
 48117 48216: gap of 100 bp
 48217 52618: contig of 4402 bp in length
 52619 52718: gap of 100 bp
 52719 56592: contig of 3874 bp in length
 56593 56692: gap of 100 bp
 56693 59635: contig of 2843 bp in length
 59636 59735: gap of 100 bp
 59736 63661: contig of 3926 bp in length
 63662 63761: gap of 100 bp
 63762 68437: contig of 4676 bp in length
 68438 68537: gap of 100 bp
 68538 71458: contig of 2921 bp in length
 71459 71558: gap of 100 bp
 71559 76888: contig of 5330 bp in length
 76889 82113: contig of 5125 bp in length
 82114 82213: gap of 100 bp
 82214 88220: contig of 6007 bp in length
 88221 88320: gap of 100 bp
 88321 93499: contig of 5179 bp in length
 93500 93599: gap of 100 bp
 93600 97901: contig of 4302 bp in length
 97902 98001: gap of 100 bp
 98002 103016: contig of 5015 bp in length
 103017 103116: gap of 100 bp
 103117 109178: contig of 6062 bp in length
 109179 109278: gap of 100 bp
 109279 117367: contig of 8029 bp in length
 117368 117407: gap of 100 bp
 117408 124079: contig of 6672 bp in length
 124080 124179: gap of 100 bp
 124180 131281: contig of 7102 bp in length


```

promoter      224..844
              /gene="ABCA1"
exon          845..1147
              /gene="ABCA1"
              /number=1

BASE COUNT   278 a  313 c   328 g   248 t

ORIGIN
Query Match      99.3%; Score 459.8; DB 9; Length 1167;
Best Local Similarity 99.8%; Pred. No. 2.5e-93;
Matches 461; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CQTGGCTGCGGGAAGCTGGAGTCTGCGGCGAGCCCGAGCCGCGCTTC 60
Db 555 CQTGGCTGCGGGAAGCTGGAGTCTGCGGCGAGCCCGAGCCGCGCTTC 614
QY 61 CCGCGGCTCTTAGCGCGGCGGCGGCGGCGGGAAGGAGCGAGCCGCGACCTAA 120
Db 615 CCGCGGCTCTTAGCGCGGCGGCGGCGGCGGGAAGGAGCGAGCCGCGACCTAA 674
QY 121 GACACCTGCTGCTACCTCCACCCGACCCACCCACCTCCGCGGAGCTAGATGCT 180
Db 675 GACACCTGCTGCTACCTCCACCCGACCCACCCACCTCCGCGGAGCTAGATGCT 734
QY 181 CGTGGCGGCTGACGCTGCGCCGCTTTAAGGGGCGGCGGCGGCTCCACGCTCTTCTGCT 240
Db 735 CGTGGCGGCTGACGCTGCGCCGCTTTAAGGGGCGGCGGCGGCTCCACGCTCTTCTGCT 794
QY 241 GAGTGACTGAACCTACATAACAGAGGCGGCGGAGCGGCGGCGGAGGAGGAGCACAGG 300
Db 795 GAGTGACTGAACCTACATAACAGAGGCGGCGGAGCGGCGGCGGAGGAGGAGCACAGG 854
QY 301 CTTTGACCGATAGTAACCTCTGCGCTCGGTCAGCCGCAATCTATATAAGGAACCTAGTCCC 360
Db 855 CTTTGACCGATAGTAACCTCTGCGCTCGGTCAGCCGCAATCTATATAAGGAACCTAGTCCC 914
QY 361 GGCAAAACCCCGCTAATTGGGAGCGAGTGTAGTGGGCGGCGGAGCCCGAGCGCGAGCC 420
Db 915 GGCAAAACCCCGCTAATTGGGAGCGAGTGTAGTGGGCGGCGGAGCCCGAGCGCGAGCC 974
QY 421 GACCCCTCTCTCCGCGGCTGGGCGAGGCGGCGGCGGAGGCTC 463
Db 975 GACCCCTCTCTCCGCGGCTGGGCGAGGCGGCGGCGGAGGCTC 1017

RESULT 7
AC021246      59570 bp   DNA      linear      HTG 13-JUL-2000
LOCUS        Homo sapiens clone RP11-IN10, LOW-PASS SEQUENCE SAMPLING.
DEFINITION   AC021246
ACCESSION   AC021246
VERSION      AC021246.2 GI:9119882
KEYWORDS     HTG: HTGS, PHASE0.
SOURCE       Homo sapiens.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 59570)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
              Anderson,S., Baldwin,J., Barna,N., Becker,L.R., Beda,F.,
              Bonuskavkiy,L., Bonkhgalter,B., Brown,A., Burkett,G., Castle,A.,
              Chepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
              Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenster,J.,
              Ferreira,P., FitzHugb,W., Forrest,C., Gage,D., Galagan,J.,
              Garcina,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
              Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
              Landers,T., Lehoczy,J., Levine,R., Lien,C., Liu,G., Locke,K.,
              Macdonald,P., Marquis,N., McEwan,P., McEwan,P., McKernan,K.,
              McPheters,R., Meldrim,J., Menus,L., Morrow,J., Naylor,J.,
              Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,

```

Pierre,N., Pisci,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Vassiliev,E., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6705871.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L2512
Center clone name: 1_N_10

* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 871: contig of 871 bp in length
872 871: gap of 100 bp
972 1834: contig of 863 bp in length
1835 1934: gap of 100 bp
1935 2804: contig of 870 bp in length
2805 2904: gap of 100 bp
2905 3745: contig of 841 bp in length
3746 3845: gap of 100 bp
3846 4696: contig of 851 bp in length
4697 4796: gap of 100 bp
4797 5640: contig of 844 bp in length
5641 5740: gap of 100 bp
5741 6540: contig of 800 bp in length
6541 6640: gap of 100 bp
6641 7509: contig of 869 bp in length
7510 7609: gap of 100 bp
7610 8479: contig of 870 bp in length
8480 8579: gap of 100 bp
8580 9430: contig of 851 bp in length
9431 9530: gap of 100 bp
9531 10376: contig of 846 bp in length
10377 10476: gap of 100 bp
10477 11322: contig of 846 bp in length
11323 11422: gap of 100 bp
11423 12302: contig of 880 bp in length
12303 12402: gap of 100 bp
12403 13280: contig of 878 bp in length
13281 13380: gap of 100 bp
13381 14241: contig of 861 bp in length
14242 14341: gap of 100 bp
14342 15196: contig of 855 bp in length
15197 15296: gap of 100 bp
15297 16123: contig of 827 bp in length
16124 16223: gap of 100 bp
16224 17072: contig of 849 bp in length
17073 17172: gap of 100 bp
17173 18041: contig of 869 bp in length
18042 18141: gap of 100 bp
18142 19009: contig of 866 bp in length
19010 19109: gap of 100 bp
19110 19966: contig of 857 bp in length


```

SOURCE          human.
ORGANISM         Homo sapiens
REFERENCE        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS          Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE            (Bases 1 to 183999)
JOURNAL          Hayden,M.R., Brooks-Wilson,A.R., Pimstone,S.N. and Cleeve,S.M.
                 Compositions and methods for modulating hcl cholesterol and
                 triglyceride levels
FEATURES         Patent: WO 0115676-A 1 08-MAR-2001;
                 University of British Columbia (CA) ; Xecon Genetics Inc. (CA)
                 Location/Qualifiers
                 source
                   1..183999
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
BASE COUNT      49549 a 37944 c 41170 g 54950 t      386 others
ORIGIN
Query Match     96.5%; Score 447; DB 6; Length 183999;
Best Local Similarity 98.7%; Pred. No. 1.1e-90;
Matches 462; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY 1 CTTTGCTCCGGGACGCTGACTAGAGAGTCTGCGGCGAGCCCGGAGCCCGGAGCCGCTTC 60
Db 28413 CTTTGCTCCGGGACGCTGACTAGAGAGTCTGCGGCGAGCCCGGAGCCCGGAGCCGCTTC 28472
QY 61 CCGCGCGTCTTAGCGCGCGCGCGCGGGGAGAGGGGAGCGAGACCGCGGACCCCTAA 120
Db 28473 CCGCGCGTCTTAGCGCGCGCGCGGGGCGGGGAGAGGGGAGCGAGACCGCGGACCCCTAA 28532
QY 121 GACACCTGCTGTACCTTCA-----CCCCACCCACCCACCTCCCGGAGAGGAGC 175
Db 28533 GACACCTGCTGTACCTTCA-----CCCCACCCACCCACCTCCCGGAGAGGAGC 28592
QY 176 TGTGCTGGGCGGCTGAACCTCTGCGGCGGCGCGCGCGCGCGCGCGCGCTCTTT 235
Db 28593 TGTGCTGGGCGGCTGAACCTCTGCGGCGGCGCGCGCGCGCGCGCGCGCTCTTT 28652
QY 236 CTGCTGAGTCACTGAACCTCTGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGAGC 295
Db 28653 CTGCTGAGTCACTGAACCTCTGCGGCGGCGCGCGCGCGCGCGCGCGCGCGAGC 28712
QY 256 ACAGGCTTTGACCGATAGTAACCTCTGCGGCGGCGCGCGCGCGCGCGCGCGCGAGC 355
Db 28713 ACAGGCTTTGACCGATAGTAACCTCTGCGGCGGCGCGCGCGCGCGCGCGCGAGC 28772
QY 356 GTCCCGGCAAAACCCCGTAATTCGAGCGAGAGTGAAGTGGGCGGCGCGCGCGCGAGCC 415
Db 28773 GTCCCGGCAAAACCCCGTAATTCGAGCGAGAGTGAAGTGGGCGGCGCGCGCGAGCC 28832
QY 415 GAGCGGACCCCTCTCTCCCGGCGTGGCGGAGGCGGCGGCGGCGGCGGAGCTC 463
Db 28833 GAGCGGACCCCTCTCTCCCGGCGTGGCGGAGGCGGCGGCGGCGGCGGAGCTC 28880

RESULT 9
AX351029 AX351029 3231 bp DNA linear PAT 06-FEB-2002
LOCUS     Sequence 1 from Patent WO0183746.
DEFINITION
ACCESSION AX351029
VERSION   AX351029.1 GI:18616385
KEYWORDS
SOURCE    human.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS   Rosier-Montus,M.F., Prades,C., Lencine,C., Naudin,L., Deneffe,P.,
          Brewer,B., Duverger,N., Remaley,A. and Santamarina-Fojo,S.
TITLE     Regulatory nucleic acid sequences of the abcl gene
JOURNAL   Patent: WO 0183746-A 1 08-NOV-2001;
          Aventis Pharma S.A. (FR)
FEATURES  Location/Qualifiers
          source
            1..3231

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT      803 a 773 c 876 g 773 t
ORIGIN
Query Match     96.4%; Score 446.4; DB 6; Length 3231;
Best Local Similarity 98.7%; Pred. No. 2.3e-90;
Matches 462; Conservative 0; Mismatches 1; Indels 5; Gaps 2;

QY 1 CTTTGCTCCCGGACGCTGGAGTCTGCGGCGGCGGAGGAGCGAGCCCGGAGCCGCTTC 60
Db 2517 CTTTGCTCCCGGACGCTGGAGTCTGCGGCGGCGGAGGAGCGAGCCCGGAGCCGCTTC 2576
QY 61 CCGCGCGTCTTAGCGCGCGCGCGGGGAGAGGGGAGCGAGACCGCGGAGCCCTAA 120
Db 2577 CCGCGCGTCTTAGCGCGCGCGGGGCGGGGAGAGGGGAGCGAGACCGCGGAGCCCTAA 2636
QY 121 GACACCTGCTGTACCTTCCAC-----CCCCACCCACCCACCTCCCGGAGAGG 175
Db 2637 GACACCTGCTGTACCTTCCAC-----CCCCACCCACCCACCTCCCGGAGAGG 2696
QY 176 TGTGCTGGGCGGCTGAACCTCTGCGGCGGCGGCGGCGGCGGCGGCTTCCAGCTTC 235
Db 2697 TGTGCTGGGCGGCTGAACCTCTGCGGCGGCGGCGGCGGCGGCGGCTTCCAGCTTC 2756
QY 236 CTGCTGAGTCACTGAACCTCTGCGGCGGCGGCGGCGGCGGCGGCGGAGGAGC 295
Db 2757 CTGCTGAGTCACTGAACCTCTGCGGCGGCGGCGGCGGCGGCGGCGGAGGAGC 2816
QY 296 ACAGGCTTTGACCGATAGTAACCTCTGCGGCGGCGGCGGCGGCGGCGGAGC 355
Db 2817 ACAGGCTTTGACCGATAGTAACCTCTGCGGCGGCGGCGGCGGCGGCGGAGC 2876
QY 356 GTCCCGGCAAAACCCCGTAATTCGAGCGAGAGTGAAGTGGGCGGCGGCGGCGGAGCC 415
Db 2877 GTCCCGGCAAAACCCCGTAATTCGAGCGAGAGTGAAGTGGGCGGCGGCGGCGGAGCC 2936
QY 415 GAGCGGACCCCTCTCTCCCGGCGTGGCGGAGGCGGCGGCGGCGGAGCTC 463
Db 2937 GAGCGGACCCCTCTCTCCCGGCGTGGCGGAGGCGGCGGCGGCGGAGCTC 2984

RESULT 10
AF275948 AF275948 149034 bp DNA linear PRI 17-JUL-2000
LOCUS     Homo sapiens ABCA1 (ABCA1) gene, complete cds.
DEFINITION
ACCESSION AF275948
VERSION   AF275948.1 GI:9247085
KEYWORDS
SOURCE    Homo sapiens.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 149034)
AUTHORS   Santamarina-Fojo,S., Peterson,K.M., Knapper,C., Qiu,Y., Freeman,L.,
          Cheng,S.F., Gorsic,J., Remaley,A., Yang,X.P., Haudenschild,C.,
          Prades,C., Chimini,G., Blackmon,E., Francois,T., Duverger,N.,
          Rubin,E.M., Rosier,M., Deneffe,P., Fredrickson,D.S. and Brewer,H.B.
          Jr.
TITLE     Complete genomic sequence of the human ABCA1 gene: analysis of the
          human and mouse ATP-binding cassette A promoter
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (14), 7987-7992 (2000)
MEDLINE   20345099
PUBMED    10884428
REFERENCE 2 (bases 1 to 149034)
AUTHORS   Santamarina-Fojo,S., Peterson,K.M., Knapper,C.L., Freeman,L.A.,
          Remaley,A.T., Yang,X.-P., Haudenschild,C.C., Blackmon,E.B.,
          Francois,T.L. and Brewer,H.B. Jr.
TITLE     Direct Submission
JOURNAL   Submitted (08-JUN-2000) Molecular Disease Branch, National
          Institutes of Health, National Heart, Lung and Blood Institute,
          Bethesda, MD 20892, USA
FEATURES  Location/Qualifiers

```

```

source
1. .149034
/organism="Homo sapiens"
/db_xref="taxon:9606"
1454. .148034
/gene="ABCA1"
mRNA
Join(1454. .1674,25831. .25989,40385. .40478,45012. .45153,
46423. .46541,67729. .67850,70831. .71007,83960. .84052,
89010. .89250,91952. .92101,92433. .92549,96758. .96955,
97702. .97907,98428. .98604,100391. .100613,102360. .102581,
103642. .103846,104951. .105064,106862. .107033,
108023. .108154,109460. .109602,109806. .109943,
110346. .110666,112124. .112196,113183. .113385,
115053. .115101,115297. .115410,116806. .116954,
118604. .118728,119964. .120062,123094. .123283,
124804. .124898,126208. .126240,127363. .127468,
128943. .129017,129539. .129708,130939. .131116,
133114. .133229,133341. .133485,134526. .134649,
135736. .135865,136130. .136250,137038. .137100,
138008. .138114,140469. .140510,140982. .141116,
142060. .142163,142646. .142738,143397. .143640,
144581. .148034)
/gene="ABCA1"
/product="ABCA1"
Join(1454. .1674,25831. .25923)
5'UTR
2406. .2678
/gene="ABCA1"
repeat_region
/rpt_family="Alu"
4240. .4509
/rpt_family="Alu"
repeat_region
4998. .5270
/rpt_family="Alu"
repeat_region
5563. .5839
/rpt_family="Alu"
repeat_region
6615. .6877
/rpt_family="Alu"
repeat_region
9360. .9046
/rpt_family="Alu"
repeat_region
9998. .10279
/rpt_family="Alu"
repeat_region
11859. .12133
/rpt_family="Alu"
repeat_region
12810. .12902
/rpt_family="Alu"
repeat_region
15220. .15399
/rpt_family="Alu"
repeat_region
15508. .15787
/rpt_family="Alu"
repeat_region
18600. .18860
/rpt_family="Alu"
repeat_region
20950. .21206
/rpt_family="Alu"
repeat_region
25038. .25314
/rpt_family="Alu"
CDS
Join(25924. .25989,40385. .40478,45012. .45153,46423. .46541,
67729. .67850,70831. .71007,83960. .84052,89010. .89250,
91962. .92101,92433. .92549,96758. .96955,97702. .97907,
98428. .98604,100391. .100613,102360. .102581,103642. .103846,
104951. .105064,106862. .107033,108023. .108154,
109460. .109602,109806. .109943,110646. .110866,
112124. .112196,113183. .113385,115053. .115101,
115297. .115410,116806. .116954,118604. .118728,
119964. .120062,123094. .123283,124804. .124898,
126208. .126240,127363. .127468,128943. .129017,
129539. .129708,130939. .131116,133114. .133229,
133341. .133485,134526. .134649,135736. .135865,
136130. .136250,137038. .137100,138008. .138114,
140469. .140610,140982. .141116,142060. .142163,
142646. .142738,143397. .143640,144581. .144721)
/gene="ABCA1"
/codon_start=1
/product="ABCA1"
/protein_id="AA886276.1"
/db_xref="GI:9247086"
/translation="MACPQQLRLKLLKKNLTPRRROTCOLLEVAWPEFTFLILLISVRL
SYPPYQHECEHPNKAMPAGTLPWQGLTICANNPECREYPTPGAGPVGVGNKSTIV
ARLFDARLLLYSKQDTSKDMRXLVRLTILOQIKSSSNLKLQDFLVDNETSGFLYH
KLSLPKSTVDMKRADVILHKVFLQGYQLHLTSLONGSKSEEMILQGDQVSELDGLP
REKLAARAEVLRSNDIILKPIRLTNSPTSPFESKELAEATKTLHLISLGTIAGELFSMR
KSWSDRQVEAMLTIVNSSSSTQIYQAVSRIVCGHPEGGLXIKSLNMYEDNNYKALF
GNGTEAEATFTFOELAVFDLEGMWELSPKINTFMENSCOEMLVRLMLDSRDNDHFE
ROMVAENKTFQELAVFDLEGMWELSPKINTFMENSCOEMLVRLMLDSRDNDHFE
QQLDGLDTAQDIAFALAKHPEDVSSNGSVYTWREAFNETINCAITISRPNECVNLN
KLEPIATEVWLINKSMELLDERKFWAGVFTGTPGSTELPHHYKIKLMDIDNVERT
NKIKDGYDMPGRADPFDMDRYWGGFAVLQDVQCAITRLVTGTEKTGVTYQOMPY
PCYVDITFLVMSRSMPLMTLAWIYSVAIVIKGYVEKARLKTMTIMEDNSILW
FSWFITSSILPILVSAGLLVILKGLNLPPYSDVVFVLSFAVVTILQCLISTLF
SRANLAAACGGIIVFTLYLPLVYLCAWODYVGTFLKIPAXLLSPVAFGFGCYFPALFE
EQCLGVQNDNLPESPVEEDGFNLTTSVSMMLFDTFLYGVWVWYIEAVPGCYGIPSPW
YFPCIKSWFGSEESDEKSHPGSNQARISICEDEEPTHLKLGVSIONLVKVRDGMKV
AVDGLALNFYEGGITSFGLHNGAGKTTMSILGTLPPTSGTAYILGKDIREMSTIR
QNLGVCQHNVLFDMLTVEEHWFYARLKGLEKEVKAEQMALDVLGIPSSLKSKT
SOLSGMQKSLVALAFVGGKSVILDEPTAGVDVPSRRGIWELLKYRQGTIILST
HMDRADYLGDRITATISHGKLCQVSSFLKNQLTGYVLTLLVKDVESSLSSCRNSS
STVSYLKXEDSVSSQSDAGLQSDHESPTLTDVSAISNLIKHYSEARLYVDIGHEL
TYLVPYEAKEGAFVELFHEIDRLSDLCISYGSSETILLEIFLKVAEESVDARETS
DGTLPARNRRAPGDKQSCRLPRTEDDAADNDSDDPESRETDLLSGDKGKGYOVK
GKMLTQQQFVALLWKRLKLRARRSRKGFQIYLPVAFVCIALVFLIYPPGKYPSLE
LQPMWYNEQYITFVSNDADEDTGTLELLNALTQDPGFGTRCMEGNTIPDTPOAGREW
TTAPVPTIMDLFQNGWNTWNPSPACQSSDKIKKMLPVCPPGAGGLPPQPKONTA
DILODLTGRTNISDYLKTYVQIIIAKSLNKIWNFRYGGISLGYVNTQALPPSQEVN
DAXKMKKHLKLAKSSADRFNLGRTMIGLDTNNVYWNKNGWRAISSEFLAVIN
NALLRANLQGENPSSHIGITAFNPLNLTQKLSSEVAMTTSVDVLVSICVIFAMSFY
PASVYFLQERSVAKAHQFISGVKPVYIWLNSFVDMCNVAVPATLIIFICFOO
KSYVSTNLPLALLLLLYGWSITPLMYPASFVFKIPSTAYVVLVSMLFINGISVA
TFVLELFTDKLNNNDILKSVLFIPLPHCLGRGLIDVKNQAMADALRFRGNFVS
P.LSWDLVGRNLFAMAVEGVVFLITVLIQYRFFIRPVPYNAKLSPLNDEDEOVRERO
RLDGGQNDILEIKELTKYRKKAKPADRCYGIFFGCEGFLIGVAGAGKSTFKM
LIGDITVTRGDAFLXANSILSNLHEVHQNMGCPOQFDATELTGREHVEFFALLRGV
PEKEKVGGEWAIRKILGYKYGKAGNYSGGNKKRLSTAMALIGGPPVVFLEDPITG
MDPKARFLWNAKLSVYKGRSVLTSMSMECEALCTRMAINVRGRCFLGVSQHLK
NRFQDGYTTVIRIAGSNPDLKPVQDFGLAFTPGSVAKERNNMLQVLPSSLSLARI
FSILSQSKRLHIEDYVSQCTILDQVVFVNFQKDSQDDHLKDLSLHKNTVVDVAULT
SFLQDEKVKESYV"
29916. .30180
/rpt_family="Alu"
31050. .31314
/rpt_family="Alu"
31692. .31973
/rpt_family="Alu"
36167. .36440
/rpt_family="Alu"
37617. .37892
/rpt_family="Alu"
37999. .38166
/rpt_family="Alu"
38172. .38269
/rpt_family="Alu"
38995. .39271
/rpt_family="Alu"
41547. .41565
/gene="ABCA1"
/ncore="A009X28"
41593. .41835
/rpt_family="Alu"
41657. .41679
/gene="ABCA1"
/ncore="A009X28"
42320. .42584
/rpt_family="Alu"
44349. .44626
/rpt_family="Alu"
48473. .48669
/rpt_family="Alu"
50547. .50614
/rpt_family="Alu"
54285. .54418

```

```

repeat_region      /rpt_family="Alu"
54849..54988
repeat_region      /rpt_family="Alu"
55029..55226
repeat_region      /rpt_family="Alu"
55297..55344
repeat_region      /rpt_family="Alu"
58501..58935
repeat_region      /note="LTR"
/rpt_family="HERV"
61246..61489
repeat_region      /rpt_family="Alu"
62812..63077
repeat_region      /rpt_family="Alu"
64082..64306
repeat_region      /rpt_family="Alu"
67123..67392
repeat_region      /rpt_family="Alu"
68499..68731
repeat_region      /rpt_family="Alu"
69481..69769
repeat_region      /rpt_family="Alu"
71398..72233
repeat_region      /rpt_family="Alu"
72244..72501

```

Query Match 96.4%; Score 446.4; DB 9; Length 145034;

Best Local Similarity 98.7%; Pred No. 15e-90; Mismatches 1; Indels 5; Gaps 1;

Matches 462; Conservative 0; Mismatches 0; Mismatches 1; Indels 5; Gaps 1;

```

QY 1 CCTTGGCTGCGGGAAGTGTGACTAGAGTCTCGGCGCAGCCCGGAGCCGCGCTTC 60
|||||
Db 1077 CCTTGGCTGCGGGAAGTGTGACTAGAGTCTCGGCGCAGCCCGGAGCCGCGCTTC 1136
QY 61 CCGGCGCTCTTAGCGCGCGCGCGCGCGGGAAGGGGACGAGCCGGGACCCCTAA 120
|||||
Db 1137 CCGGCGCTCTTAGCGCGCGCGCGCGGGAAGGGGACGAGCCGGGACCCCTAA 1196
QY 121 GACACCTCTCTACCTCTCCAC----CCCAACCCACCACTCCCTCCCACTCCCTAGA 175
|||||
Db 1197 GACACCTCTCTACCTCTCCACCCCAACCCCACTCCCTCCCACTCCCTAGA 1256
QY 176 TGTGTCGTGGCGGTGAACCTGCGCCGCTTAAAGGGCGCGCGCGCGCGCTCCACGTGCTTT 235
|||||
Db 1257 TGTGTCGTGGCGGTGAACCTGCGCCGCTTAAAGGGCGCGCGCGCGCTCCACGTGCTTT 1316
QY 236 CTGCTGAGTGACTGACTACATAACAGAGCCCGGGAAGCGGGCGGGAGGAGGAGACC 295
|||||
Db 1317 CTGCTGAGTGACTGACTACATAACAGAGCCCGGGAAGCGGGCGGGAGGAGGAGACC 1376
QY 296 ACAGGCTTTGACCGATAGTAACCTCTGCGCTCGGTCGAGCGGAATCTATAAAGGAACCTA 355
|||||
Db 1377 ACAGGCTTTGACCGATAGTAACCTCTGCGCTCGGTCGAGCGGAATCTATAAAGGAACCTA 1436
QY 356 GTCCCGGGAACCCCGTAATTGGAGCGAGAGTGTAGTGGGCGCGGACCCCGAGACC 415
|||||
Db 1437 GTCCCGGGAACCCCGTAATTGGAGCGAGAGTGTAGTGGGCGCGGACCCCGAGACC 1496
QY 416 GAGCGGACCTCTCTCCCGGCTCGGCGAGGAGCGGCGGAGCTC 463
|||||
Db 1497 GAGCGGACCTCTCTCCCGGCTCGGCGAGGAGCGGCGGAGCTC 1544

```

```

RESULT 11
AF287262      201144 bp  DNA  linear  PRI 29-APR-2001
LOCUS
DEFINITION  Homo sapiens ATP-binding cassette 1 sub-family A member 1 (ABCA1)
ACCESSION  AF287262
VERSION    AF287262.1  GI:13876612
KEYWORDS   .
SOURCE     Homo sapiens.
ORGANISM  Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 201144)
Qiu,Y., Cavelier,L., Chiu,S., Yang,X., Rubin,E. and Cheng,J.F.
Human and mouse ABCA1 comparative sequencing and transgenesis
studies revealing novel regulatory sequences
Genomics 73 (1), 56-76 (2001)
21251004
MEDLINE
PUBMED
11352567
2 (bases 1 to 201144)
Qiu,Y., Cavelier,L., Chiu,S., Rubin,E. and Cheng,J.F.
Direct Submission
Submitted (13-JUL-2000) Genome Science Department, Lawrence
Berkeley National Laboratory, 1 Cyclotron Rd, MS 84-171, Berkeley,
CA 94720, USA
FEATURES
Location/Qualifiers
1..201144
/organism="Homo sapiens"
/db_xref="taxon:9606"
33931..181457
/gene="ABCA1"
join(33931..34151,58319..58476,72893..72986,77520..77661,
78331..79649,101048..101169,104152..104328,117287..117379,
12329..122569,125285..125424,125754..125870,
130079..130276,131023..131228,131749..131925,
133708..133930,135677..135898,136959..137163,
138268..138381,140179..140350,141340..141471,
142777..142919,143123..143260,143963..144183,
145441..145513,146657..146859,148527..148575,
148771..148884,150280..150428,152078..152202,
153438..153536,156368..156757,158278..158372,
159682..159714,160837..160942,162417..162491,
163013..163182,164413..164590,166588..16703,
165815..166959,168505..168628,169715..169844,
170109..170229,171017..171079,171987..172093,
17448..174589,174961..175095,176039..176142,
176825..176717,177376..177619,178560..181457)
/gene="ABCA1"
/product="ATP-binding cassette 1 sub-family A member 1"
33931..34151
/gene="ABCA1"
/number=1
58319..58476
/gene="ABCA1"
/number=2
join(58411..58476,72893..72986,77520..77661,78931..79049,
101048..101169,104152..104328,117287..117379,
12329..122569,125285..125424,125754..125870,
130079..130276,131023..131228,131749..131925,
133708..133930,135677..135898,136959..137163,
138268..138381,140179..140350,141340..141471,
142777..142919,143123..143260,143963..144183,
145441..145513,146657..146859,148527..148575,
148771..148884,150280..150428,152078..152202,
153438..153536,156368..156757,158278..158372,
159682..159714,160837..160942,162417..162491,
163013..163182,164413..164590,166588..16703,
166815..166959,168505..168628,169715..169844,
170109..170229,171017..171079,171987..172093,
17448..174589,174961..175095,176039..176142,
176825..176717,177376..177619,178560..178700)
/gene="ABCA1"
/codon_start=1
/product="ATP-binding cassette 1 sub-family A member 1"
/protein_id="AAK43526.1"
/db_xref="GI:13876613"
/translation="MACPQLRLLLKNNLFRRTQCTQLLEVAWPLFIFLILISVRL
SYPEQHECFPNKRSAGTLFWOGIICNANPCRPYTPGACPVGNKNSIV
ARLPSDARLLLYSKQDTSMKMRKRLTQQIKSSNKLQDFVNDVTFSSFLYH
NLSPKSTVDKMLRADYILHKVFLQGYQLHTSLCNGSKSEMIQLQDQVEYSLGLP
KEKLAARERVLRSNMDIKPLRTLNSTSPFSLAPATKILLSLGTLAQELFSMR
SKSDMRQEVNFTNVNSSSTOIVQAVSRIVCGHPGGGLKIKLSLWYEDNNYKALP
GGNGLEDAETFYDNTTPYCNLNMKNLSSPLSLRIWALKPLLVGKILVTPDTPAT

```


ROVMAEVNKTFOELAVFHDLGEMWELSPKTIWTFMENSOEMDLVRMLLDSRDHFWE
COLDC-DNTAQGLVAFIAKEDVEDOSSNGSVYTBREAFNE"NOA1R3"SRMECVNLN
KLEPATIEWLNKSNELLDERKFWAGVFTGIPGSIHELPHVAYKIRMDINDVERT
NKIDGYWDPGRADPFEDRMVWGCFAYLQDVVEQALIRVLTGFEKKTGYMQQMPY
PCYVDIFLRVNSRMPFLMTLAWISVAVIIGIVIEKEARLRETRIMGLDNLISLW
FSWFISLIPILVYAGLLVILKGLNLLPSPDPSVYFVLSFAVVTILOCFLLISLIF
SRANLAACGGIIFETLVLPPYVLCVWDQVYGTFLKIPASLLSPVAFGFGCECFALFE
EOGICVQWQDLPFESVEEDGPNLTTSVSMMLPNTFLYGVMTWYIEAVFPQGYIPRPW
VPPCKSVFEGESDEKSEPOSNOKRISEICMEEPETHLKLGVSONLVKVRDMKV
AVDGLALPNYEQIISFLHNGAGKTTMSILGLFPPTSGTAYILGKDIRSEMSIIR
QNLGVPQNVLFUDMTVEEHFWYARLKLKSKHVKAEQEMADLVGLPSPKLSKT
SOLSGMQRKSVAILAFVGGKCVWILDEPTAGVDYISRGIIWELLKYRQGTILST
HMDADVLGDRITAIISHGKLCVGSJFLKQLGTGYLTIVKRDVSSLSSCRNS
SVYLVKKPDSVSSDAGLGSDBESDTLTDVSAISNLEIRKHVSEARLVEDIGHEL
TVLYPEAKGAFVELHEIDRLSDLGSSVGETTLEBIFILKVAEESVDQETS
DGTLPARNRARGKQCLRPTEDDAADPNDSIDPESRETDLLSGMDGSGYOVK
GWLTCQGFVALLKRLIARSRKQFTAQIVLPAVFCVIALVFSLLVPPFKYPSLE
LQPMWNYEQIETVEDAPEDTGTLELLNALTQDPGFTRCMEGNPIPTDPCOAGBEEW
TTAPVPQTIMDFQNGWNTMQNPSPACOSSDKIKMLPVPFGAGGLPPQPKQNTA
DILQDIGNISDYLVKTYVQIIAKSLXNKIWNFRYGGFSLGVSNQALPPSQFVN
DAIKOMKHLKAKSDADRFLNSLRPMIGLDRNNKVMWNNKGWHAISSEFLAVIN
NALRANLOKGNPSHYGITAFNHPNLTKOOLSEVALMTTSVDVLVSCIVFPAKSFV
PASVYVFLQBRVSRAKELQFISGVKPVYIWSNFWDMCNVVPATVILITIPQOQ
KSYSTNLPVLJALLILYGSITPMPASVFKIPSTAYVYVLTISVNLFTIGINSVA
TFVLELTDNKLNNIDILKSVLFIPEHPCFLGRLIDMKVKNQAMADALERGENRVS
PLSWDLGNLFAMAVEGVYFFLITVLIQYREFIRPVPYNAKLSPLNDEDEVRERQ
RLDGGQNDILEIKELTIVYRKRKPAVDICVIGIPGCEGFLGVNGAGKSYFKM
LTGDTWVSGDAFLAKNSILSNHEVHONWGYCPOPDATTEILTGREHVEFTAILRGV
PEKEVGKGEWAIKRLGLVKYGEKYAGNYSGGNKKLSTAMALIGPPVYVDEPTIG
MDPKARFLWNCALSVMKGRSVLITSHMECEALCTRMAMVNGFRCLGSLGVOLK
NREGDGYTIVRIAGSNPDLKPVQDFGLAFPGVLKEXHRMLQYQLPSSLSLARI
FSLQSKKRLHIEDYVSQTTLDQVYVNAFAKQSDDDHLKDLISLHKNOTVVDVAVLI
SFLQDEKVESV"
72893 72986
/gene="ABCA1"
/number=3
77520..77661
/gene="ABCA1"
/number=4
76931..79049
/gene="ABCA1"
/number=5
101048..101169
/gene="ABCA1"
/number=6
104152..104328
/gene="ABCA1"
/number=7
117287..117379
/gene="ABCA1"
/number=8
122329..122569
/gene="ABCA1"
/number=9
125285..125424
/gene="ABCA1"
/number=10
125754..125870
/gene="ABCA1"
/number=11
130079..130276
/gene="ABCA1"
/number=12
131023..131228
/gene="ABCA1"
/number=13
131749..131925
/gene="ABCA1"
/number=14
133708..133930
/gene="ABCA1"
/number=15
135677..135899

/gene="ABCA1"
/number=16
136959..137163
/gene="ABCA1"
/number=17
138268..138381
/gene="ABCA1"
/number=18
140179..140350
/gene="ABCA1"
/number=19
14340..141471
/gene="ABCA1"
/number=20
142777..142919
/gene="ABCA1"
/number=21
143123..143260
/gene="ABCA1"
/number=22
143963..144183
/gene="ABCA1"
/number=23
145441..145513
/gene="ABCA1"
/number=24
146657..146859
/gene="ABCA1"
/number=25
148527..148575
/gene="ABCA1"
/number=26
148771..148884
/gene="ABCA1"
/number=27
150280..150428
/gene="ABCA1"
/number=28
152078..152202
/gene="ABCA1"
/number=29
153438..153536
/gene="ABCA1"
/number=30
156568..156757
/gene="ABCA1"

Query Match 95.4%; Score 441.6; DB 9; Length 201144;
Best Local Similarity 98.1%; Pred. No. 1.8e-89;
Matches 459; Conservative 0; Mismatches 4; Indels 5; Gaps 1;

QY 1 CCTTGGCTGCGGGAAAGTGGACTAGAGAGTCTGCGGCGAGCCCGAGCCGACGCGTTC 60
DB 33554 CCTTGGCTGCGGGAAAGTGGACTAGAGAGTCTGCGGCGAGCCCGAGCCGACGCGTTC 33613
QY 61 CCGCGCTCTTTAGCCCGCGGCGCCCGGGGGGAAAGGACGACGACGCGGGACCCCTAA 120
DB 33614 CCGCGCTCTTTAGCTGGCGGCGCCCGGGGGGAAAGGACGACGACGCGGGACCCCTAA 33673
QY 121 GACACCTGCTGTACCTCCAC----CCCCACCCACCCACCTCCCGCCCACTCCCTAGA 175
DB 33674 GACACCTGCTGTACCTCCACCCACCCACCCACCTCCCGCCCACTCCCTAGA 33733
QY 176 TGTCTCTGCGCGCTGAAGCTGCGCCGCTTTAAGGGGCGGCGCCCGGCTCCACGTGCTTT 235
DB 33734 TGTCTCTGCGCGCTGAAGCTGCGCCGCTTTAAGGGGCGGCGCCCGGCTCCACGTGCTTT 33793
QY 236 CTGCTGAGTACTGAACCTACATAAACAGAGCCCGGGAACCGGGGCGGGAGGAGGAGC 295
DB 33794 CTGCTGAGTACTGAACCTACATAAACAGAGCCCGGGAACCGGGGCGGGAGGAGGAGC 33853
QY 296 ACAGGCTTGACCATAGTACTCTGGGCTGGGAGCGCGGAGTCTATATAAGAACTA 355

Db 33854 ACAGCGTTTGACGGATAGTAAACCTCTCGCTGTCGAGCCGCAATCTATATAAGGAACCTA 33913
Qy 356 GTCCGGCAGAAACCCGTAATTCGCGAGCGAGTGAAGTGGCGCGGACCCGAGAGCC 415
|||||
Db 33914 GTCTCGCAAAACCCGTAATTCGCGAGCGAGTGAAGTGGCGCGGACCCGAGAGCC 33973
Qy 416 GAGCGGACCCCTCTCTCCCGGGCTGGCGGAGGCGAGGCGGGAGGCTC 463
|||||
Db 33974 GAGCGGACCCCTCTCTCCCGGGCTGGCGGAGGCGAGGCGGGAGGCTC 34021
|||||
RESULT 12
AX351031 AX351031 2893 bp DNA linear PAT 06-FEB-2002
LOCUS Sequence 3 from Patent W00183746.
ACCESSION AX351031
VERSION AX351031.1 GI:18616387
KEYWORDS
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 Rosier-Montus, M.F., Prades, C., Lemoine, C., Naudin, L., Benefie, P.,
Stewer, B., Duverger, N., Remaley, A. and Santamarina-Fojo, S.
Regulatory nucleic acid sequences of the abcl gene
Patent: WO 0183746-A 3 08-NOV-2001;
Aventis Pharma S.A. (FR)
FEATURES
Source 1..2893
Location/Qualifiers
Query Match 76.88; Score 355.4; DB 6; Length 2893;
Sect Local Similarity 98.48; Pred. No. 6.6e-70;
Matches 371; Conservative 0; Mismatches 1; Indels 5; Gaps 1;
Qy 1 CTTTGGCTGCTCGGGAACCTGGACTAGAGTCTGGCGGCGAGCCCGAGCCAGCGCTTC 60
Db 2517 CTTTGGCTGCTCGGGAACCTGGACTAGAGTCTGGCGGCGAGCCCGAGCCAGCGCTTC 2576
Qy 61 CCGGCGCTTTAGCGCGCGCGCGCGCGCGCGCGGAGGAGCGAGACCGCGGAGCCCTAA 120
Db 2577 CCGGCGCTTTAGCGCGCGCGCGCGCGCGCGGGAAGGAGCGAGACCGCGGAGCCCTAA 2636
Qy 121 GACACTGCTGTACCTCCACCCCGAC-----CCACCCACCTCCCGCCCACTCCCTAGA 175
Db 2637 GACACTGCTGTACCTCCACCCCGACCCCGACCCACCTCCCGCCCACTCCCTAGA 2696
Qy 176 TGTGTCGTGGGCGGTGAACCTGCGCCCGTTTAAGGGCGCGCGCGCGCTCCACGTCCTT 235
Db 2697 TGTGTCGTGGGCGGTGAACCTGCGCCCGTTTAAGGGCGCGCGCGCGCTCCACGTCCTT 2756
Qy 236 TGTGTCGTGACTGACTACATACATACAGCGCGGAGCGCGGCGGAGGAGGAGGAGC 295
Db 2757 TGTGTCGTGACTGACTACATACATACAGCGCGGAGCGCGGAGGAGGAGGAGGAGC 2816
Qy 296 ACAGGCTTTGACCGATAGTAACTCTGCGCTCGGCTGCGAGCCGCAATCTATAAAGGAACCTA 355
Db 2817 ACAGGCTTTGACCGATAGTAACTCTGCGCTCGGCTGCGAGCCGCAATCTATAAAGGAACCTA 2876
Qy 356 GTCCCGGCAAAAACCCC 372
Db 2877 GTCCCGGCAAAAACCCC 2893
RESULT 13
AC021246/c
LOCUS Homo sapiens clone R211-1N10, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC021246
ACCESSION AC021246

AC021246.2 GI:91119882
HTG: HTGS_PEA50.
KEYWORDS Homo sapiens.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 69570)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, B.
TITLE Homo sapiens chromosome, clone RP11-1N10
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 69570)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, B., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Bohaglavsky, L., Bouckgalter, B., Brown, A., Burkett, G., Castle, A.,
Choelev, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeArallano, K., Dewar, K., Domino, M., Doyle, M., Fenesor, J.,
Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Johnson, R., Jones, C., Kann, L., Karates, A., Klein, J.,
Landers, T., Lebacky, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Oliver, J.M., Peterson, K.,
Pierce, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.C.,
Zimmer, A. and Zody, M.
DIRECT SUBMISSION
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6705871.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://p.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2512
Center clone name: L_N10

* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 871: contig of 871 bp in length
* 372 971: gap of 100 bp
* 972 1834: contig of 863 bp in length
* 1835 1934: gap of 100 bp
* 1935 2804: contig of 870 bp in length
* 2805 2904: gap of 100 bp
* 2905 3745: contig of 841 bp in length
* 3746 3845: gap of 100 bp
* 3846 4696: contig of 851 bp in length
* 4697 4796: gap of 100 bp
* 4797 5640: contig of 844 bp in length
* 5641 5740: gap of 100 bp
* 5741 6540: contig of 800 bp in length
* 6541 6640: gap of 100 bp
* 6641 7509: contig of 866 bp in length
* 7510 7609: gap of 100 bp
* 7610 8479: contig of 870 bp in length
* 8480 8579: gap of 100 bp


```

|||||
Db 67402 AGCGAGTGTAGTGGCGGCGGACCGACAGCAGCGACGACCTCTCTCCGGCTGTC 67343
|||||
QY 442 GGCGAGGCGAGGCGCGGAGCTC 463
|||||
Db 67342 GACAGGACAGGCGGCGGAGCTC 67321

RESULT 14
AC021345/c
LOCUS
DEFINITION Homo sapiens clone RP11-24J9, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC021345
VERSION AC021345.2 GI:9130845
KEYWORDS HTG; HTGS_PHASE0.
SOURCE
ORGANISM Homo sapiens.
Homo sapiens
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 90698)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Boguslavsky, I., Boukhvalter, B., Brown, A., Burkett, G., Castle, A.,
Choquel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeArrellano, K., Dewar, K., Domino, M., Doyle, M., Fenesstor, J.,
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Leccocky, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrum, J., Menius, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Vlei, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6705761.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1483
Center clone name: 24_J_9
-----
* NOTE: This record contains 92 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 910: contig of 910 bp in length
* 911 1010: gap of 100 bp
* 1011 1873: contig of 863 bp in length
* 1874 1973: gap of 100 bp
* 1974 2824: contig of 851 bp in length

```

```

2825 2924: gap of 100 bp
2925 3802: contig of 878 bp in length
3803 3902: gap of 100 bp
3903 4816: contig of 914 bp in length
4817 4916: gap of 100 bp
4917 5759: contig of 843 bp in length
5760 5859: gap of 100 bp
5860 6764: contig of 905 bp in length
6765 6864: gap of 100 bp
6865 7747: contig of 883 bp in length
7748 7847: gap of 100 bp
7848 8753: contig of 908 bp in length
8756 8855: gap of 100 bp
8856 9753: contig of 898 bp in length
9754 9853: gap of 100 bp
9854 10757: contig of 904 bp in length
10758 10857: gap of 100 bp
10858 11732: contig of 875 bp in length
11733 11832: gap of 100 bp
11833 12739: contig of 907 bp in length
12740 12839: gap of 100 bp
12840 13710: contig of 871 bp in length
13711 13810: gap of 100 bp
13811 14684: contig of 874 bp in length
14685 14784: gap of 100 bp
14785 15662: contig of 878 bp in length
15663 15762: gap of 100 bp
15763 16677: contig of 915 bp in length
16678 16777: gap of 100 bp
16778 17676: contig of 901 bp in length
17679 17778: gap of 100 bp
17779 18679: contig of 901 bp in length
18680 18779: gap of 100 bp
18780 19632: contig of 853 bp in length
19633 19732: gap of 100 bp
19733 20634: contig of 902 bp in length
20635 20734: gap of 100 bp
20735 21620: contig of 886 bp in length
21621 21720: gap of 100 bp
21721 22573: contig of 859 bp in length
22580 22679: gap of 100 bp
22680 23568: contig of 889 bp in length
23569 23668: gap of 100 bp
23669 24554: contig of 886 bp in length
24555 24654: gap of 100 bp
24655 25521: contig of 867 bp in length
25522 25621: gap of 100 bp
25622 26487: contig of 866 bp in length
26488 26587: gap of 100 bp
26588 27464: contig of 877 bp in length
27465 27564: gap of 100 bp
27565 28466: contig of 902 bp in length
28467 28566: gap of 100 bp
28567 29464: contig of 898 bp in length
29465 29564: gap of 100 bp
29565 30447: contig of 883 bp in length
30448 30547: gap of 100 bp
30548 31453: contig of 906 bp in length
31454 31553: gap of 100 bp
31554 32452: contig of 899 bp in length
32453 32552: gap of 100 bp
32553 33447: contig of 895 bp in length
33448 33547: gap of 100 bp
33548 34435: contig of 888 bp in length
34436 34535: gap of 100 bp
34536 35433: contig of 898 bp in length
35434 35533: gap of 100 bp
35534 36440: contig of 907 bp in length
36441 36540: gap of 100 bp
36541 37422: contig of 882 bp in length
37423 37522: gap of 100 bp
37523 38402: contig of 880 bp in length
38403 38502: gap of 100 bp

```

```
* 38503 39380: contig of 878 bp in length
* 3938- 35480: gap of 100 bp
* 39481 40407: contig of 927 bp in length
* 40408 40507: gap of 100 bp
* 40508 41405: contig of 898 bp in length
* 41406 41505: gap of 100 bp
* 41506 42399: contig of 894 bp in length
* 42400 42499: gap of 100 bp
* 42500 43357: contig of 858 bp in length
* 43358 43457: gap of 100 bp
* 43458 44356: contig of 899 bp in length
* 44357 44456: gap of 100 bp
* 44457 45325: contig of 869 bp in length
* 45326 45425: gap of 100 bp
* 45426 46305: contig of 880 bp in length
* 46306 46405: gap of 100 bp
* 46406 47302: contig of 897 bp in length
* 47303 47402: gap of 100 bp
* 47403 48293: contig of 891 bp in length
* 48294 48393: gap of 100 bp
* 48394 49280: contig of 887 bp in length
* 49281 49380: gap of 100 bp
* 49381 50250: contig of 870 bp in length
* 50251 50350: gap of 100 bp
* 50351 51209: contig of 859 bp in length
* 51210 51309: gap of 100 bp
* 51310 52221: contig of 912 bp in length
* 52222 52321: gap of 100 bp
* 52322 53205: contig of 884 bp in length
* 53206 53305: gap of 100 bp
* 53306 54225: contig of 920 bp in length
* 54226 54325: gap of 100 bp
* 54326 55205: contig of 881 bp in length
* 55207 55306: gap of 100 bp
* 55307 56165: contig of 859 bp in length
* 56166 56265: gap of 100 bp
* 56266 57124: contig of 859 bp in length
* 57125 57224: gap of 100 bp
* 57225 58121: contig of 897 bp in length
* 58122 58221: gap of 100 bp
* 58222 59075: contig of 854 bp in length
* 59076 59175: gap of 100 bp
* 59176 60058: contig of 883 bp in length
* 60059 60158: gap of 100 bp
* 60159 61067: contig of 909 bp in length
* 61068 61167: gap of 100 bp
* 61168 62051: contig of 884 bp in length
* 62052 62151: gap of 100 bp
* 62152 63022: contig of 871 bp in length
* 63023 63122: gap of 100 bp
* 63123 64026: contig of 904 bp in length
* 64027 64126: gap of 100 bp
* 64127 65015: contig of 889 bp in length
* 65016 65115: gap of 100 bp
* 65116 65975: contig of 860 bp in length
* 65976 66075: gap of 100 bp
* 66076 66954: contig of 879 bp in length
* 66955 67054: gap of 100 bp
* 67055 67947: contig of 893 bp in length
* 67948 68047: gap of 100 bp
* 68048 68934: contig of 887 bp in length
* 68935 69034: gap of 100 bp
* 69035 69910: contig of 876 bp in length
* 69911 70010: gap of 100 bp
```

Query Match 49.6% Score 229.8; DB 2; Length 90698;
Best Local Similarity 56.0%; Pred. No. 7.8e-42;
Matches 279; Conservative 0; Mismatches 141; Indels 3; Gaps 3;

QY 41 AGCCCGGAGCCACGGCTCCCGCGCGCTCTAGCGCGCGCGCGCGCGCGCGCGGGAAGGG 100
|||||

Db 87653 AGCCNN 87604

```
QY 101 ACGCAGACGGCGGACCTAAGACACDCTGCTGTAGCTTCACCCCAACCCACCCACCTCC 160
Db 87803 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 57744
QY 161 CCCCACTCCCTAGATGCTGCTGGCGGCTGAAAGTCCCGCTTTAAGGGGGGGCC 220
Db 87743 CCACCTCCCGCTATAGTGTGCTGCGCTGAAGCTCGNCCG-TTAAAGGGGGGGCC 87685
QY 221 GGTCCACCGTCTTCTGCTGAGTGAACCTACATAACAGAGCGCGGGAACGGGGC 280
Db 87684 GGTCCACCGTCTTCTGCTGAGTGAACCTACATAACAGAGCGCGGGAACGGGGC 87626
QY 281 GGGAGAGGGGAGAGACAGAGCTTGGACGATAGTAACCTCTGCGTCTGGTGGAGCGGAAT 340
Db 87625 GGGAGAGGGGAGAGACAGAGCTTGGACGATAGTAACCTCTGCGTCTGGTGGAGCGGAAT 87566
QY 341 CTATAAAGGAACIAGTCCGGGCAAAACCCGTAATTCGGACGAGAGTGAGTGGGGC 400
Db 87565 CTATAAAGGAACIAGTCCGGGCAAAACCCGTAATTCGGACGAGAGTGAGTGGGGC 87507
QY 401 GGGACCGCGAGAGCGGACCGCTTCTCCCGGCTGGCGAGCGGCGGGGGAG 460
Db 87506 GGGACCGCGAGAGCGGACCGCTTCTCCCGGCTGGCGAGCGGCGGGGGAG 87447
QY 461 CTC 463
Db 87446 CTC 87444

RESULT 15
AF258627 Homo sapiens ATP binding cassette transporter 1 (ABCA1) mRNA, partial cds. 697 bp mRNA linear PRI 11-MAY-2000
DEFINITION AF258627
ACCESSION AF258627
VERSION AF258627.1 GI:7769707
KEYWORDS
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C., Acuzerat,B.E., Fielding,C.C. and Kane,J.P.
TITLE Analysis of hABC1 gene 5' end: additional peptide sequence, promoter region, and four polymorphisms
JOURNAL Biochem. Biophys. Res. Commun. 271 (2000) In press
REFERENCE 2 (bases 1 to 697)
AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C., Acuzerat,B.E., Fielding,C.C. and Kane,J.P.
TITLE Direct Submission
JOURNAL Submitted (19-APR-2000) Cardiovascular Research Institute, University of California, San Francisco, 505 Parnassus Avenue, San Francisco, CA 94143-0130, USA
FEATURES
source Location/Qualifiers
1..697
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q31"
/tissue_type="placenta"
1..>697
/gene="ABCA1"
/note="ABC1"
396..>697
/gene="ABCA1"
/note="membrane-bound"
/codon_start=1
/product="ATP binding cassette transporter 1"
/protein_id="AA569513.1"
/db_xref="GI:7769708"
/translation="MACWQFLRLLLKWLFRPQTCQLLLEVAWPLFLILLISYRL  
SYPPYEQHECHFPKAMPKSAGTLPVQVGLICNNANPCERYTPGEAPGVVGNFKS"
```

BASE COUNT 152 a 198 c 190 g 156 t 1 others
ORIGIN

Query Match 37.4%; Score 173; DB 9; Length 697;
Best Local Similarity 100.0%; Pred. Nc. 7.7e-29;
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 291 AGAGCACAGGCTTTGACCGATAGTACCTCTGGCTGGCTGCAGCCGAATCTATAAAGG 350
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 AGAGCACAGGCTTTGACCGATAGTACCTCTGGCTGGCTGCAGCCGAATCTATAAAGG 60
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 351 AACTAGTCCCGGCAAAACCCCGTAATTGCGAGCGAGAGTGAAGTGGCGCGGACCCGCA 410
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 AACTAGTCCCGGCAAAACCCCGTAATTGCGAGCGAGAGTGAAGTGGCGCGGACCCGCA 120
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 411 GAGCCGAGCGACCTTCTCTCCGGGCTCGCGGAGGGCAGGGGGGAGGTC 463
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 GAGCCGAGCGACCTTCTCTCCGGGCTCGCGGAGGGCAGGGGGGAGGTC 173
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

Search completed: February 4, 2003, 01:26:54
Job time : 1675.59 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 16:39:45 ; Search time 876.075 Seconds
(without alignments)
8559.21C Million cell updates/sec

Title: US-09-596-141c-3_COPY_1181_1643

Perfect score: 463

Sequence: 1 ccttgctgcgggaacgtg.....caggcgaggcgaggagctc 463

Scoring table: IDENTITY_NJC
Gapop 10.0 , Gapext 1.0

Searched: 16154056 seqs, 805774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*

```

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfn.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_man.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rnd.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89.4	19.3	736	9	AU135588
2	89	19.2	547	9	AL698654
3	89	19.2	763	9	AU121731
4	75	16.2	292	14	244377
5	75	15.2	1004	11	BC034824
6	60.2	13.0	998	12	BG678861

ALIGNMENTS

RESULT 1
AU135588
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AU135588 PLACE1 Homo sapiens cDNA clone PLAC1002437 5', mRNA
sequence.
AU135588
AU135588.1 GI:10396127
EST.
human.
Homo sapiens
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Iscogai,T.

ERI human cDNA Project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986

Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
Research Institute; cDNA library construction; Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.

AL105991 Drosophila
AL296898 Tetraodon
AL108460 Drosophila
AL053013 Drosophila
AL064271 Drosophila
AL098882 Drosophila
AL107514 Drosophila
AL108336 Drosophila
AL059666 Drosophila
AL184922 Tetraodon
AC081217 Pan trogl
BQ53013 Drosophila
BQ943816 AGENCOURT
AG073529 Pan trogl
BQ944204 AGENCOURT
AL514015 AL514015
BQ53024 AGENCOURT
AL554951 AL554951
AG157500 Pan trogl
BM560246 AGENCOURT
AG041031 Pan trogl
BQ219067 AGENCOURT
AG043475 Pan trogl
BQ086652 H3128H07-
AG043475 Pan trogl
BQ892481 AGENCOURT
AG065289 Pan trogl
AL513813 AL513813
AL071777 Drosophila
AL309889 Tetraodon
AL066742 Drosophila
AG042878 Pan trogl
BQ680479 AGENCOURT
AL513741 AL513741
AG060229 Pan trogl
BF262458 HV_GEA000

```

FEATURES
source
Location/Qualifiers
1..736
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PLACE1002437"
/clone_lib="PLACE1"
/tissue_type="placenta"
/notes="vector: pME18SPL3"
BASE COUNT 163 a 199 c 199 g 170 t 5 others
ORIGIN

Query Match 19.3%; Score 89.4; DB 9; Length 736;
Best Local Similarity 98.9%; Pred. No. 1.2e-11;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 373 GTAATTCGAGCGAGTGTAGTGGCGCGGAGCCGCGAGCGCGGAGCTC 432
Db 1 GTAGTTCGAGCGAGTGTAGTGGCGCGGAGCCGCGAGCGCGGAGCTC 60
|||||
QY 433 CCGGCTCGCGAGCGAGCGCGGAGCTC 463
Db 61 CCGGCTCGCGAGCGCGGAGCTC 91
|||||

RESULT 2
LOCUS AL698654 547 bp mRNA linear EST 21-MAR-2002
DEFINITION DKF2p686N12109_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
ACCESSION AL698654
VERSION AL698654.1 GI:19619194
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 547)
AUTHORS Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann
,S.
TITLE EST (Duesterhoeft, et al.)
JOURNAL Unpublished (1995)
COMMENT Contact: Duesterhoeft A
MIPS Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKF2); Email s.wiemann@dkfz-heidelberg.de;
consortium of the German Genome Project.
No sl sequence available.
This clone (DKF2p686N12109) is available at the R2PD in Berlin.
Please contact the R2PD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin- Charlottenburg, GERMANY; Email: clone@r2pd.de.
Location/Qualifiers
1..547
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKF2p686N12109"
/clone_lib="686 (synonym: hlcc3)"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/notes="vector: pRiplex2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"
BASE COUNT 112 a 154 c 157 g 124 t
ORIGIN

Query Match 19.2%; Score 89; DB 9; Length 547;
Best Local Similarity 100.0%; Pred. No. 1.4e-11;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 AATTCGAGCGAGTGTAGTGGCGCGGAGCCGCGAGCGCGGAGCTC 434
|||||

FEATURES
source
Location/Qualifiers
1..763
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MAMMAL000851"
/clone_lib="MAMMAL"
/tissue_type="mammary gland"
/notes="vector: pME18SFL3"
BASE COUNT 137 a 205 c 260 g 158 t 3 others
ORIGIN

Query Match 19.2%; Score 89; DB 9; Length 763;
Best Local Similarity 100.0%; Pred. No. 1.5e-11;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 AATTCGAGCGAGTGTAGTGGCGCGGAGCCGCGAGCGCGGAGCTC 434
|||||
Db 1 AATTCGAGCGAGTGTAGTGGCGCGGAGCCGCGAGCGCGGAGCTC 60
|||||

RESULT 4
LOCUS 244377 292 bp mRNA linear EST 14-NOV-1994
DEFINITION HSC12B081 normalized infant brain cDNA Homo sapiens cDNA clone
c-12b08, mRNA sequence.
ACCESSION 244377
VERSION 244377.1 GI:573506
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 292)

```


AUTHORS

Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Boulgatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Plesu, G., Poulliot, I., Sebastiani-Kabakchis, C., and Tessier, A.

TITLE

IMAGE: molecular integration of the analysis of the human genome and its expression

JOURNAL

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

MEDLINE

95277534

COMMENT

Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read

Genexpress_library_id: C; Genexpress_sequence_id: y-c-1zb08
Seq primer: (-21)M13-universal.

FEATURES

source

1..292
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-1zb08"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site: HindIII; Site: NotI; sex: female; dev_stage=3 months old; isolate=muscular atrophy patient; tissue_type=total brain; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector. Clone library from B.Souares, Psychiatry Dept. Columbia University, USA. Normalization_method: Bento Soares, P.N.A.S in press"

BASE COUNT 50 a 87 c 96 g 56 t 3 others
ORIGIN
Query Match 16.2%; Score 75; DB 14; Length 292;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 389 GTAGTGGGCGGGACCCGAGCGAGCGAGCCGACCGCTTCTCTCCGGGCTGGCGCAGGG 448
|||||
Db 1 GTAGTGGGCGGGACCCGAGCGAGCGAGCCGACCGCTTCTCTCCGGGCTGGCGCAGGG 60
QY 449 CAGGGCGGGGAGCTC 463
|||||
Db 61 CAGGGCGGGGAGCTC 75

RESULT 5

BC034824

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REMARK

CONTACT

EMAIL

TISSUE

PROJECT

URL

CONTACT

EMAIL

TISSUE

PROJECT

URL

CONTACT

EMAIL

TISSUE

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amc@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 52 Row: d Column: 19
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5915657
This clone has the following problem: incomplete processing.

Location/Qualifiers
1..1004
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4749735"
/tissue_type="Skin, squamous cell carcinoma"
/clone_lib="NCI_CGAP_Skn4"
/lab_host="DH103"
/note="Vector: pCMV-SPORT6.cdb"

FEATURES

source

1..1004
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4749735"
/tissue_type="Skin, squamous cell carcinoma"
/clone_lib="NCI_CGAP_Skn4"
/lab_host="DH103"
/note="Vector: pCMV-SPORT6.cdb"

BASE COUNT 306 a 234 c 230 g 234 t
ORIGIN
Query Match 16.2%; Score 75; DB 11; Length 1004;
Best Local Similarity 100.0%; Pred. No. 3.7e-08;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 389 GTAGTGGGCGGGACCCGAGCGAGCGAGCCGACCGCTTCTCTCCGGGCTGGCGCAGGG 448
|||||
Db 1 GTAGTGGGCGGGACCCGAGCGAGCGAGCCGACCGCTTCTCTCCGGGCTGGCGCAGGG 60
QY 449 CAGGGCGGGGAGCTC 463
|||||
Db 61 CAGGGCGGGGAGCTC 75

RESULT 6
BG578861
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BC678861 998 bp mRNA linear EST 01-MAY-2001
602624750F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4749735 5',
mRNA sequence.
BG678861
BG678861.1 GI:13910258
EST.
human.
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC <http://mgi.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bcm.tmc.edu
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: InCyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM0503 row: g column: 16
High quality sequence stop: 860.

Location/Qualifiers
1..998
/organism="Homo sapiens"
/db_xref="taxon:9606"

[illegible]

```
QY 181 CCGGGCGGTGAACGTG 199
|||||
Db 263 GGGGGGGGGGGGGCGG 281

RESULT 9
CNS017SV
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN37J08 of Drosophila library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION AL108460.1 GI:5628764
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 1101)
AUTHORS
TITLES
JOURNAL
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN37J08"
/clone_lib="DrosBAC"
/plasmid="pBelOBAC11"
/note="end : SP6"
BASE COUNT 254 a 176 c 160 g 152 t 359 others
ORIGIN
Query Match 10.9%; Score 50.6; DB 17; Length 1101;
Best Local Similarity 13.4%; Pred. No. 0.032;
Matches 53; Conservative 187; Mismatches 157; Indels 0; Gaps 0;

QY 28 GAGTCGCGCGGAGCCGAGCCGCGCTTCCGCGGCTTTAGCGCGGGGGCGG 87
|||||
Db 575 GKKKGGVGRVCCAGGAGCACMAADGGCCAKWACCSSSSSSGSCASTSSAS 734

QY 88 CGCGGGAAGGGACCGACCGGAGCCCTAGACACCTGCTGTACCTCCACCCCCAC 147
|||||
Db 735 RGVWSSCAGSGGAGSAGGAGGGGGGGGCGCCASSVCCAGSASCSASMC 784

QY 148 CCACCCACCTCCCGCACTCCCTAGATGTGCTGGGGGCTGAACGTCCCGTTA 207
|||||
Db 795 GVSSGSSCSAGCCGVSASVSSASVSKVASAVASVAGSAGAVSSCR 854

QY 208 AGGGGCGGCGCGGCTCCAGCTTCTGCTGAGTGAAGTGAACATAAACAGAGC 267
|||||
Db 855 SVASVSAASVSSVSSSSSVSAASVSSASSSASNAVAASVAVSSVAS 914

QY 268 CGGAACGGGGGAGGAGGAGCAGCAGGCTTTCACCGATAGTAACCTTCGCGCTC 327
|||||
Db 915 WSSSSSSSSSVVSVASVAAASVSSSSSSSVSTSSASVSVSVSVSVSS 974

QY 328 GTGTCAGCGCAATCTATAAAGGAACCTAGTCCCGGCAAAACCCCGTAATTCGCGAGGAG 387
|||||
Db 975 ASSSSSVSVSVVAVASAAAAAASSSASAVAVVSSSSSSSSSSSSSS 1034

QY 388 AGTGAAGTGGGGGAGCCGAGCCGAGCGGAGCGGAGCC 424
|||||
Db 1035 SVSSVSSSCSVSVSSSVSVAVASASASVSV 1071

RESULT 10
CNS0091P/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence IET3 end of BAC #
BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
VERSION AL053013.1 GI:4934451
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 925)
AUTHORS
TITLES
JOURNAL
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org the BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoler in Peter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
BcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
1..925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RPCI-98"
/note="end : IET3"
BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN
Query Match 10.4%; Score 48.2; DB 17; Length 925;
Best Local Similarity 14.3%; Pred. No. 0.12;
Matches 34; Conservative 116; Mismatches 87; Indels 0; Gaps 0;

QY 16 ACCTGGACTAGAGTGTGGCGGAGCCGAGCCGAGCCCTAACACACTGCTGAC 75
|||||
Db 779 ASSSSASSSRSRSGGAGGAGGAGSSSSSSSSASAGSVSSASSSSSSSSV 720

QY 76 CGCGGCGCGCGGGGAGGGAGCGGAGCGGAGCCGAGCCCTAACACACTGCTGAC 135
|||||
Db 719 ASSMSCSBSBSASASSSSSSSSSSASGACSCCCTSWSCSTASMSAARSSSSSS 660

QY 136 CTCACCCCGCCACCCACCTCCCGCCCACTCCCTAGATGTGCTGGGGGCTGAAC 195
|||||
Db 659 SSSSMASSSASSSSSSSSSSSSSSSSSGACGBSMSSGGGSGSVSSASSMS 600

QY 196 GTCCGCGCGTTAAAGCGCGGCGCGCGCTCCAGTGTCTTCTGCTGAGTGAAC 252
|||||
Db 599 RSSGSGGGGCGGSGSSSSSGSGSGSGSVSCSSGCMCRCSAAASCAVAS 543

RESULT 11
CNS006EZ/c
```

LOCUS CNS006EZ 870 bp DNA linear GSS 03-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC # BAC13E15 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL0364271
 VERSION AL064271.1 GI:4944346
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 870)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/Drosophila_bac.htm.

FEATURES
 source
 1..870
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="BAC13E15"
 /clone="RPCI-98"
 /notes="end : TET3"
 BASE COUNT 256 a 90 c 127 t 310 others
 ORIGIN
 Query Match 10.3%; Score 47.6; DB 17; Length 870;
 Best Local Similarity 27.3%; Pred. No. 0.17;
 Matches 63; Conservative 73; Mismatches 95; Indels 0; Gaps 0;
 QY 2 CTTCGCTCCGGGACGTCGACTAGACAGTCGTGGCGGCGGAGGAGCGGACGCGGACGCGCTTCC 61
 Db 301 CTSSCTTTGGCKSSGGGTTSSCBTBTGCGSCSTSSSTCTCTCTTCGGSCTSYSS 742
 QY 62 CGCGCGTCTTAGCGGCGGCGGCGGCGGAGGAGGAGCGGACGCGGACGCGCTTAAG 121
 Db 741 SSTTCBCSSSGSGSKANRSGSGGSSYVVGASWSGCGKASAGKGRKASSGHCCS 682
 QY 122 ACACCTGCTGTACCTCCACCCACCCACCCACCCACCTCCGCCCACTCCCTCCAGTAGTC 181
 Db 581 SCGSGSGSGSYACCCGCCSCYCCSCYCCSCYCCSCYCCSCYCCSCYCCSCYCCSCYCC 622
 QY 182 GTGGCGGCTGACGTCGCGCTTTAAGGCGGCGGCGGCGGCGGCTCCAGTCC 232
 Db 621 MGSSEGGNGTGCYCGCTCTCTCCSYCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 571
 RESULT 12
 CNS010EW 1009 bp DNA linear GSS 26-JUL-1999
 LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
 DEFINITION BACN03p19 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL098882
 VERSION AL098882.1 GI:5610493
 KEYWORDS GSS.

LOCUS CNS01720 1101 bp DNA linear GSS 26-JUL-1999
 LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
 DEFINITION BACN17N07 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL107514
 VERSION AL107514.1 GI:5627725
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) -

SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1003)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billand at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES
 Location/Qualifiers
 1..1009
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACN03P19"
 /clone_lib="DrosBAC"
 /plasmid="pBelobAC11"
 /notes="end : SP6"
 BASE COUNT 147 a 377 c 178 g 64 t 243 others
 ORIGIN
 Query Match 10.3%; Score 47.6; DB 17; Length 1009;
 Best Local Similarity 37.3%; Pred. No. 0.17;
 Matches 85; Conservative 50; Mismatches 88; Indels 5; Gaps 1;
 QY 6 GCTGCGGAGACGTGACTAGACAGTCGTGGCGGCGGACGCGGACGCGCTTCCGGG 65
 Db 858 SCCNCGRCGCGCGCGGACGCGSSGSSGSSGSSGSSGCGCGCGCGCGCGGGG 799
 QY 66 CGTCITTAGCGCGCGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 125
 Db 798 SGSSSSCGGACCGGCGNSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 739
 QY 126 CTGTGTACCTTCACCCACCCACCCACCCACCTCCGCCCACTCCCTCCAGTAGTCGTGG 185
 Db 738 CSSTGTGCGGCG 684
 QY 186 GCGGCTGACGTCGCGCGITTAAGGCGGCGGCGCGCGCTCCAGTGGT 233
 Db 683 GGGGKGGMGCGGCGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGGSGG 636
 RESULT 13
 CNS01720/c
 LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
 DEFINITION BACN17N07 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL107514
 VERSION AL107514.1 GI:5627725
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) -

Search completed: February 4, 2003, 03:26:33
Job time : 888.789 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 12:56:40 ; Search time 21.7309 Seconds
(without alignments)
6534.066 Million cell updates/sec

Title: US-09-596-141c-3_COPY_1181_1643

Perfect score: 463

Sequence: 1 cctggctgcgggaacgtg.....cagggcagggcgaggagtc 463

Scoring table: IDENTITY_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 441362 seqs, 1533381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2.6/ptodata/1/ina/5A_COMB.seq.*

2: /cgn2.6/ptodata/1/ina/5B_COMB.seq.*

3: /cgn2.6/ptodata/1/ina/6A_COMB.seq.*

4: /cgn2.6/ptodata/1/ina/6B_COMB.seq.*

5: /cgn2.6/ptodata/1/ina/PCTUS_COMB.seq.*

6: /cgn2.6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38.6	8.3	4481	US-03-041-886-18	Sequence 18, Appl
2	38.4	8.3	595	US-08-483-533-4	Sequence 4, Appl
3	38.4	8.3	595	US-09-283-471A-4	Sequence 4, Appl
4	38.4	8.3	1327	US-08-483-533-36	Sequence 36, Appl
5	38.4	8.3	1327	US-09-283-471A-36	Sequence 36, Appl
6	38.4	8.3	1335	PCT-US91-06532-1	Sequence 1, Appl
7	37.6	8.1	936	US-08-018-977C-4	Sequence 4, Appl
8	36.2	7.8	43280	US-08-804-227C-1	Sequence 1, Appl
9	35.8	7.7	4085	US-08-072-574-7	Sequence 7, Appl
10	35.8	7.7	4181	US-08-072-574-9	Sequence 9, Appl
11	35.9	7.7	11907	US-08-061-376-4	Sequence 4, Appl
12	35.6	7.7	4411529	US-09-103-840A-1	Sequence 1, Appl
13	35.6	7.7	4411529	US-09-103-840A-1	Sequence 1, Appl
14	35.2	7.6	35100	US-08-306-694B-19	Sequence 19, Appl
15	35.2	7.5	35100	PCT-US93-05251-19	Sequence 19, Appl
16	35	7.6	7218	US-08-232-463-14	Sequence 14, Appl
17	34.6	7.5	1280	US-08-483-533-38	Sequence 38, Appl
18	34.6	7.5	1280	US-09-283-471A-38	Sequence 38, Appl
19	34.6	7.5	17410	US-07-841-646-3	Sequence 3, Appl
20	34.6	7.5	17410	US-08-147-023-3	Sequence 3, Appl
21	34.6	7.5	17410	US-08-447-570-3	Sequence 3, Appl
22	34.6	7.5	17410	US-08-449-700-3	Sequence 3, Appl
23	34.6	7.5	17410	US-08-449-699A-3	Sequence 3, Appl
24	34.4	7.4	954	US-08-170-294-1	Sequence 1, Appl
25	34.4	7.4	954	US-08-664-855-1	Sequence 1, Appl
26	34.4	7.4	954	US-09-049-289-1	Sequence 1, Appl
27	34.4	7.4	2888	US-08-765-907A-1	Sequence 1, Appl

C 28	34.2	7.4	1490	4	US-08-964-127-3	Sequence 3, Appl
C 29	34.2	7.4	1490	4	US-09-496-692-3	Sequence 3, Appl
C 30	34.2	7.4	1610	4	US-09-276-531-56	Sequence 56, Appl
C 31	34.2	7.4	2460	4	US-08-964-127-1	Sequence 1, Appl
C 32	34.2	7.4	2460	4	US-09-496-692-1	Sequence 1, Appl
C 33	34.2	7.4	4085	1	US-08-486-270-7	Sequence 7, Appl
C 34	34.2	7.4	4085	3	US-08-367-264-7	Sequence 7, Appl
C 35	34.2	7.4	4085	4	US-09-153-757-7	Sequence 7, Appl
C 36	34.2	7.4	4181	1	US-08-486-270-9	Sequence 9, Appl
C 37	34.2	7.4	4181	3	US-08-367-264-9	Sequence 9, Appl
C 38	34.2	7.4	4181	4	US-09-153-757-9	Sequence 9, Appl
C 39	34.2	7.4	4207	4	US-08-660-148-1	Sequence 1, Appl
C 40	34.2	7.4	4207	4	US-08-660-148-3	Sequence 3, Appl
C 41	34.2	7.4	4303	4	US-08-660-148-4	Sequence 4, Appl
C 42	34.2	7.4	4303	4	US-08-660-148-6	Sequence 6, Appl
C 43	34.2	7.4	72604	4	US-09-268-992-7	Sequence 7, Appl
C 44	34.2	7.4	72604	4	US-09-657-474-7	Sequence 7, Appl
C 45	33.8	7.3	2877	4	US-09-235-103-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-09-041-886-18
; Sequence 18, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sherroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4481 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 163..4099
US-09-041-886-18

Query Match 8.3%; Score 38.6; DB 4; Length 4481;

Best Local Similarity 52.1%; Pred. No. 0.43; Indels 0; Gaps 0;

Matches 86; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

CY 7 CTGCCGGAACTGTGACTAGAGASTCTGCGGCGCGAGCCCGAGCCGCTTCCGCGC 66

Db 194 CCGGTTGGACAGGCTCGCGCTTCGCGCAGCCAGGTGGCGGCTCGC 253
Qy 57 GTCTTAGCGCGCGCGCGCGGGAAGGGAGCGAGACCGGAGCCCTAACACACC 126
Db 254 TCCAGCGCGCGCGCGGAGCGCGGCGGCGGCGGCTCGCGCGCGCGCGGACCGTATC 313
Qy 127 TCGCTAGCTCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCC 171
Db 314 CTTCCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 358

RESULT 2

US-08-483-533-4
; Sequence 4, Application US/0848333
; Patent No. 6172047
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; APPLICANT: Chou, Joany
; TITLE OF INVENTION: Method for Treating Tumorigenic Diseases
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,533
; FILING DATE: 07-MAR-95
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/419,853
; FILING DATE: 11-APR-95
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/861,253
; FILING DATE: 31-MAR-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, James P.
; REGISTRATION NUMBER: 28,491
; REFERENCE/DOCKET NUMBER: 28097/32742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-483-533-4

Query Match 8.3%; Score 38.4; DB 4; Length 595;
Best Local Similarity 49.6%; Pred. No. 0.34;
Matches 127; Conservative 0; Mismatches 126; Indels 3; Gaps 1;

Qy 1 CCTTGGCTGCGCGGAGACCTGGAGCTAGAGACTCTCGGGGCGAGCCCGGAGCGGCTC 60
Db 109 CTTGCGCGCGCTGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 168
Qy 61 CCGCGCGCTTAGCG 120
Db 169 CCGCGGACCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 228

Qy 121 GACACCTCTGTACCTCCACCCCGCCACCC---CACCCACCTCCCGCCACCTCCCTAGATG 177
Db 229 GACCCCG 288
Qy 178 TGTCTGGCG 237
Db 289 CCACCTGTGGTGTCTGGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 348
Qy 238 GCTGAGTGAATGAAT 253
Db 349 GCGGCGCGCGCGCGGCT 364

RESULT 3

US-09-283-471A-4
; Sequence 4, Application US/09283471A
; Patent No. 6340673
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; APPLICANT: Chou, Joany
; TITLE OF INVENTION: Method For Treating Tumorigenic Diseases
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/283,471A
; FILING DATE: 04-APR-1999
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/851,233
; FILING DATE: 31-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/419,853
; FILING DATE: 11-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/483,533
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, James P.
; REGISTRATION NUMBER: 28,491
; REFERENCE/DOCKET NUMBER: 27373/32742A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-283-471A-4

Query Match 8.3%; Score 38.4; DB 4; Length 595;
Best Local Similarity 49.6%; Pred. No. 0.34;
Matches 127; Conservative 0; Mismatches 126; Indels 3; Gaps 1;

Qy 1 CCTTGGCTGCGCGGAGACCTGGAGCTAGAGACTCTCGGGGCGAGCCCGGAGCGGCTC 60
Db 109 CTTGCGCGCGCTGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 168
Qy 61 CCGCGCGCTTAGCG 120

Db 550 CTGCGGCGCTCGCTGCGAGCGCGGCGGCGGAGGCGGCGGAGCGGCGGCGGAG 609
QY 61 CCGCGGCTGTAGCGCGGCGGCGGCGGCGGAGGAGAGCGGAGCGGCGGAGCGGCTAA 120
Db 610 CCGCGGAGCGGCGGAGCGGCGGCGGCGGCGGAGCGGCGGCGGAGCGGCGGCGGAG 659
QY 121 GACACCTGCTGTAGCTTCCACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGAG 177
Db 670 GACCGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAG 723
QY 178 TGTGCTGGGCGGCTGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 237
Db 730 CGACCTGGTGTGTGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTCG 789
QY 238 GCTGAGTGTAGTGAAT 253
Db 790 GCGGCGGCGGCGGCT 805

RESULT 6
PCT-US91-06532-1
; Sequence 1, Application PC/TUS9106532
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; TITLE OF INVENTION: Recombinant Herpes Simplex Viruses
; TITLE OF INVENTION: Vaccines and Methods
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, C'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: Two First National Plaza Suite 2100
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/06532
; FILING DATE: 199-09-10
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gruber, Lewis S.
; REGISTRATION NUMBER: 30,060
; REFERENCE/DOCKET NUMBER: 27373/8235
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/346-5750
; TELEFAX: 312/394-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1335 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
PCT-US91-06532-1

Query Match 8.3%; Score 38.4; DB 5; Length 1335;
Best Local Similarity 49.6%; Pred. No. 0.39;
Matches 127; Conservative 0; Mismatches 126; Indels 3; Gaps 1;
QY 1 CTTTGGCTGGCGGAGCTGTAGAGTCTCGGCGGAGCGGCGGCGGCGGCGGCGGCTC 60
Db 558 CTTGCGCGCTGTGGCTGGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 617
QY 61 CCGCGCGCTGTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
Db 618 CCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 677

QY 121 GACACCTGCTGTAGCTTCCACCGCGGCGGCGGCGGCGGCGGCGGCGGAGTATG 177
Db 678 GACCGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 737
QY 178 TGTGCTGGGCGGCTGACGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 237
Db 738 CGACCTGGTGTGTGCGGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTCG 797
QY 238 GCTGAGTGTAGTGAAT 253
Db 798 GCGGCGGCGGCGGCT 813

RESULT 7

US-08-018-977C-4/c
; Sequence 4, Application US/08018977C
; Patent No. 5686601
; GENERAL INFORMATION:
; APPLICANT: Weber, Peter C.
; TITLE OF INVENTION: DNA Sequences Encoding Mutant Antiviral
; TITLE OF INVENTION: Regulatory Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Pennsylvania State University
; STREET: 113 Technology Center
; CITY: University Park
; STATE: PA
; COUNTRY: US
; ZIP: 16802-7000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS 6.22
; SOFTWARE: Microsoft Word for Windows 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/018,977C
; FILING DATE: 18-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/726,071
; FILING DATE: 05-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas J. Monahan
; REGISTRATION NUMBER: 29,835
; REFERENCE/DOCKET NUMBER: 91-1039/2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (814) 865-6277
; TELEFAX: (814) 865-3591
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 936 bases
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-018-977C-4

Query Match 8.1%; Score 37.6; DB 1; Length 936;
Best Local Similarity 54.3%; Pred. No. 0.58;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 32 CTGCGGCGGAGCGGCGGCGGCGGCTTCCGCGGCGGCTTAGCGCGGCGGCGGCGGCG 91
Db 785 CTGCGGAGCGGCGGCGGCGGCTTCTGCTCGACAGAGCGGCGGCGGCGGCGGCGGCG 726
QY 92 GGGAGGAGGAGCGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 151
Db 725 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 656
QY 152 CCCACCTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 171
Db 665 CGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 646


```
Query Match 7.7% Score 35.6; DB 4; Length 4411529;
Best Local Similarity 54.2%; Pred. No. 8.2;
Matches 77; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 40 CAGCCCGGAGCCAGCCGCTCCCGCGGTCTTAGCCGCGCGGCGCGCGCGGCGGGAAGG 99
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1630671 CAGCCGCGCAGCCGCGCGCGGTCTTAGCCGCGCGGCGCGCGGCGGCGGAGGC 1630730

QY 100 GACCGAGCCGCGACCTAGACACCTGCTGACCTCCACCCGCCACCCACCCACCTC 159
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1630731 GCGCGAGCCGCGGAGACCAAGAGGAGTGTGAAGCCGCGCCACCCGCTGCCGCCCGT 1630790

QY 160 GCCCAACTCCCTAGATGTGTC 181
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1630791 GGCCACCGGTCGCCACCGGTGCC 1630812

RESULT 13
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 7.7% Score 35.6; DB 4; Length 4411529;
Best Local Similarity 55.7%; Pred. No. 8.2;
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 43 CCCCAGCCAGCCGCTCCCGCGGTCTTAGCCGCGGCGCGCGCGCGGCGGGAAGGGAC 102
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 836449 CCCCAGCCAGCCGCTAGCGGCGGCTAGCCCGCGCGCGCGCGCGCGCGCGCACCA 836390

QY 103 GCAGACCGGGACCTAGACACCTGCTGACCTCCACCCCGCGCGCGCGCGCGCGCGCG 162
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 836389 ACAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 836330

QY 163 CC 164
|||
DB 836329 CC 836328

RESULT 14
US-08-651B-15/c
; Sequence 19, Application US/08305691B
; Patent No. 5734039
; GENERAL INFORMATION:
; APPLICANT: Calabretta, Bruno
; APPLICANT: Skorski, Tomasz
; TITLE OF INVENTION: ANTISENSE
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seidel, Gonda, Lavorina & Monaco, P.C.
; STREET: Two Penn Center, Suite 1800
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: U.S.A.
```

```
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720 kb
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/306.691B
FILING DATE: September 15, 1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A.
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8321-8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
TELEX: No. 5734039e
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 35100 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-306-6913-19

Query Match 7.6% Score 35.2; DB 1; Length 35100;
Best Local Similarity 50.0%; Pred. No. 4.5;
Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 57 CTTCGCCGCGCTTAGCCGCGCGCGCGCGCGGGAAGGGAGCGAGACCGCGGACC 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8471 CCGTCCCGGACACCTTGTCAGTCTGCCCATAGATGATCTGGACACCTTAGGGGAG 8412

QY 117 CTAGACACCTGCTGACCTCCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 176
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8411 AAGTGAAGAAAGAGGACATCCACTGCCCGCGCGCGCGCGCGCGCGCGCGCGCG 8352

QY 177 GTGTGCGGCGCGCTGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 232
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 8351 ATGTGCGCGCGCGCTCTGTGTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8296

RESULT 15
PCI-US93-06251-19/c
; Sequence 15, Application PC/TUS9306251
; GENERAL INFORMATION:
; APPLICANT: Wickstrom, Eric and Rife, Jason P.
; TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing
; TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: NY
; COUNTRY: USA
; ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCI/US93/06251
FILING DATE: 19930630
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,346
```

```
; REFERENCE/DOCKET NUMBER: 8586
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FCT-US93-08251.19

Query Match: 7.6%; Score 35.2; DB 5; Length 35100;
Best Local Similarity 50.0%; Pred. No. 4.5;
Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 57 CTTCCTCCGCGCTCTTAGCGCGGCGGCGGCGGGAAGGGACGCGAGACCGCGGAC 116
DB 8471 COTGCCCGGACACCTGTGTGATGATGATGATGATGATGATGATGATGATGAT 8412
QY 117 CTAAGACACCTGCTGTACCTCCACCCGACCCGACCTCCGCGGCGGCGGCGGCGG 176
DB 8411 AAGTGAAGAGAGAGAGATCCACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8352
QY 177 GTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 232
DB 8351 ATGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 8296
```

Search completed: February 3, 2003, 21:01:05
Job time : 7799.73 secs


```
;
;
; FOR PREPARING ARTIFICIAL CHROMOSOMES
;
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe LLP
; STREET: 4350 La Jolla Village Drive, 7th Floor
; CITY: San Diego
; STATE: CA
; COUNTRY: USA
; ZIP: 92122
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/125,767
; FILING DATE: 17-Apr-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/724,693
; FILING DATE: 28-NOV-2000
; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1995
; APPLICATION NUMBER: 08/629,822
; FILING DATE: 10-APR-1996
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24601-402J
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8403
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
;
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42999 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; MOLECULE TYPE: Genomic DNA
;
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 17:
;
; US-10-125-767-17
;
;
; Query Match 8.0%; Score 37.2; DB 9; Length 42999;
; Best Local Similarity 51.9%; Pred. No. 1.1;
; Matches 84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
;
; QY 14 GAAGTGGATAGAGATCTGCGCGGAGCCGAGCCGCGCTCCCGGGGCTTTAG 73
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
; Db 10013 GAGAGTGAAGGAGCGCGGGCGCGCGCGCGCGCGGTGTGTGCTGGAG 10072
;
; QY 74 GCCGCGGCGCGCGCGGGGAGGAGGAGCCAGACCCGCGGACCTAGACACCTGCTGTA 133
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
; Db 10073 GCGCGCGCGCGCGCGGGGAGGAGGAGTGTGGGTCCTCCCGCGCGCGCGCGCGCC 10132
;
; QY 134 CCCTCCACCCACCCACCCACCCACCTCCCGCCAACTCCCTAGA 175
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
; Db 10133 TCTTCCCTCCCTCCCGCGCGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCG 10174
;
;
; RESULT 8
; US-09-772-304A-1/c
; Sequence 1, Application US/09772304A
; Patent No. US20020146785A1
; GENERAL INFORMATION:
; APPLICANT: Manishi, L.H.
; TITLE OF INVENTION: NOVEL ESCHERICHIA COLI HAVING ACCESSION
; FILE REFERENCE: A33943 066123.0103
; CURRENT APPLICATION NUMBER: US/09/772,304A
; CURRENT FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4826
; TYPE: DNA
; ORGANISM: Streptomyces aureofaciens
;
; US-09-772-304A-1
;
; Query Match 7.9%; Score 36.6; DB 10; Length 4826;
; Best Local Similarity 49.7%; Pred. No. 0.91;
; Matches 93; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
;
; QY 42 GCGCGAGCGCCACCGCTTCCCGCGGTCTTAGCGCGGGCGCGCGCGGGGAAGGGGA 101
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
; Db 947 GCGCGCGCCACCGCGGAGCACACCGGGCGCGCGCGCTCCGACCGAGCGGTGCG 888
;
; QY 102 CGCAGACCGCGGACCCCTAAGACACCTGCTGTACCTCCACCCACCCACCCACCTCC 161
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
; Db 887 ACCCGCGCGGACCGGAGCCGCTCCGCGGTCCCGCGCGACAGCTCCGACAGCACGTCA 828
;
; QY 162 CCAACTCCCTAGATGTGTGCGGGGCGGTGAACGTGCGCGCTTTAAGGGGCGGGCCCG 221
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
; Db 827 CCGCGCGCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 768
;
; QY 222 GCTCCAC 228
; |||||
;
; Db 767 CCGCGCG 761
;
;
; RESULT 9
; US-09-833-381-1386
; Sequence 1386, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NO. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1386
; LENGTH: 1102
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(1102)
; OTHER INFORMATION: n = A,T,C or G
;
; US-09-833-381-1386
;
; Query Match 7.8%; Score 36.2; DB 10; Length 1102;
; Best Local Similarity 51.2%; Pred. No. 0.8;
; Matches 109; Conservative 0; Mismatches 103; Indels 1; Gaps 1;
;
; QY 123 CACTGCTGTACCTCCACCCCGCCACCCACCTCCCGCCAACTCCCTAGATGTGTCG 182
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
; Db 665 CACTCCAGTCTGTGTCAGCTTACCTCAGCCACCATCTGGGCACTGCTAGTAAAGCG 724
;
; QY 183 TGGCGCGCTGACGTCCCGCGTTAAGGGGCGCGCGCGCGCTCCACGCTTCTGCTGA 242
; ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
;
; Db 725 TCAGGGGATCCAGGTCCCGCGGTATGCGGAGACGCCAAAGCT-CAAGAGCAGTCGGCAGA 783
```


Tue Feb 4 09:39:45 2003

us-09-596-141c-3_copy_1181_1643.rnpb

Page 7

Search completed: February 3, 2003, 16:31:11
Job time : 95.1333 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 16:39:45 ; Search time 666.044 Seconds
(without alignments)
8559.210 Million cell updates/sec

Title: us-09-596-141c-3_copy_1292_1643

Perfect score: 352
Sequence: 1 ggcacctagaacacctgtgtg.....cagggcagggcgaggagctc 352

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32306132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estau.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_estl.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_othe.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89.4	25.4	736	9 AUI35588	AUI35588 AUI35588
2	89	25.3	547	9 AL698654	AL698654 DKF25686N
3	89	25.3	763	9 AUI21731	AUI21731 AUI21731
4	75	21.3	292	14 Z44377	Z44377 HSC12B081 n
5	75	21.3	1004	11 BC034824	BC034824 Homo sapi
6	60.2	17.1	998	12 BG678861	BG678861 602624760

C 7	47.6	13.5	925	17	CNS0091P	AL053013 Drosophil
C 8	43	12.2	935	17	CNS006XK	AL066051 Drosophil
C 9	42.4	12.0	910	17	CNS006GN	AL065629 Drosophil
C 10	42.2	12.0	925	17	CNS0091P	AL053013 Drosophil
C 11	39.6	11.2	912	17	CNS006N3	AL065775 Drosophil
C 12	39.4	11.2	1101	17	CNS00397	AL063912 Drosophil
C 13	38.8	11.0	384	17	CNS035G7	AL228688 Tetradon
C 14	38.8	11.0	827	17	CNS0137-	AL102515 Drosophil
C 15	38.8	11.0	844	17	CNS0052P	AL056652 Drosophil
C 16	38.8	11.0	885	13	BM415333	BM415333 Op20407 M
C 17	38.4	10.9	490	9	AL580352	AL580352 A-580352
C 18	38.4	10.9	1101	17	CNS017S-	AL108460 Drosophil
C 19	38.4	10.9	1257	14	BQ675446	BQ675446 AGENCOURT
C 20	37.8	10.7	908	13	BM416197	BM416197 Op21284 M
C 21	37.8	10.7	1863	17	GGA200110	AJ231825 Gallus ga
C 22	37.2	10.6	1177	14	BQ677491	BQ677491 AGENCOURT
C 23	37.2	10.6	1201	17	CNS016BR	AL106545 Drosophil
C 24	36.8	10.5	840	12	BF265554	BF265554 HV_EA001
C 25	36.8	10.5	1101	17	CNS012JV	AL101653 Drosophil
C 26	36.6	10.4	881	17	CNS00512	AL065702 Drosophil
C 27	36.6	10.4	1203	17	CNS015V4	AL106054 Drosophil
C 28	36.4	10.3	1201	17	CNS007K1	AL067365 Drosophil
C 29	36.2	10.3	833	14	BQ219067	BO219067 AGENCOURT
C 30	36.2	10.3	933	17	AG073525	AG073525 Pan trogl
C 31	36	10.2	384	17	CNS043G6	AL272751 Tetradon
C 32	36	10.2	740	14	BQ178338	BQ178338 UT-M-ERO-
C 33	36	10.2	966	17	CNS06SUF	AL413725 17 end of
C 34	36	10.2	1101	17	CNS00GP3	AL072367 Drosophil
C 35	36	10.2	1409	12	BF303894	BF303894 601886514
C 36	36	10.2	1446	17	AG072447	AG072447 Pan trogl
C 37	35.8	10.2	487	17	CNS0060P	AL062040 Drosophil
C 38	35.8	10.2	1011	17	CNS051KG	AL316969 Tetradon
C 39	35.8	10.2	1100	14	BM921360	BM921360 AGENCOURT
C 40	35.8	10.2	1101	17	CNS00EPO	AL069493 Drosophil
C 41	35.8	10.2	1201	17	CNS015VQ	AL106075 Drosophil
C 42	35.6	10.1	513	10	BEG18114	BEG18114 601462521
C 43	35.6	10.1	1101	17	CNS01720	AL107514 Drosophil
C 44	35.4	10.1	809	17	AG136883	AG136883 Pan trogl
C 45	35.4	10.1	857	17	BH157466	BH157466 ENTRP347F

ALIGNMENTS

RESULT 1
AUI35588
LOCUS AUI35588 PLACE1 Homo sapiens CDNA cline PLACE1002437 5', mRNA
DEFINITION AUI35588 linear EST 02-AUG-2002
ACCESSION AUI35588
VERSION AUI35588.1 GI:10996127
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 736)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagei,T., Sugano,S. and Isogai,T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0612, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5' - 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

```

FEATURES
  source
    Location/Qualifiers
      1. .736
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="PLACE:002437"
        /clone_lib="2ACEL"
        /tissue_type="placenta"
        /note="vector: pME185FL3"
BASE COUNT      163 a   199 c   199 g   170 t      5 others
ORIGIN

Query Match      25.4%; Score 83.4; DB 9; Length 736;
Best Local Similarity 98.9%; Pred. No. 1.6e-13;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 262 GTAATTCGAGCGAGTGTAGTGGCGCGGACCCGAGAGCGGAGCCCTTCTCTC 321
|||||
DB 1 GTAGTGGCAGCGAGTGTAGTGGCGCGGACCCGAGAGCGGAGCCCTTCTCTC 60
|||||

QY 322 CCGGCTCGCGCAGCGAGCGGCGGGGAGCTC 352
|||||
DB 61 CCGGCTCGCGCAGCGAGCGGCGGGGAGCTC 91

RESULT 2
AL698654
LOCUS      AL698654      547 bp      mRNA      linear      EST 21-MAR-2002
DEFINITION DKEZp686N12109_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
ACCESSION  AL598654
VERSION     AL598654
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 547)
AUTHORS    Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann
S.
TITLE      EST (Duesterhoeft, et al.)
JOURNAL    Unpublished (1999)
COMMENT    Contact: Duesterhoeft A
MIPS
Am Klopferspitze 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@kfz-heidelberg.de;
sequenced by Qiagen (Hilden/Germany); within the cDNA sequencing
consortium of the German Genome Project.
No si sequence available.
This clone (DKEZp686N12109) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
  1. .547
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="DKEZp686N12109"
    /clone_lib="586 (synonym: hlcc3)"
    /dev_stage="adult"
    /tissue_type="human skeletal muscle"
    /lab_host="DH102"
    /note="vector: pTriplex2; Site_1: SfIIA; Site_2: SfiIB;
    cDNA-collection"
BASE COUNT      112 a   154 c   157 g   124 t
ORIGIN

Query Match      25.3%; Score 89; DB 9; Length 547;
Best Local Similarity 100.0%; Pred. No. 2e-13;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 AATTCGAGCGAGTGTAGTGGCGCGGACCCGAGAGCGGAGCCCTTCTCTCC 323
|||||
DB 1 AATTCGAGCGAGTGTAGTGGCGCGGACCCGAGAGCGGAGCCCTTCTCTCC 60
|||||

RESULT 3
AL698654
LOCUS      AL698654      763 bp      mRNA      linear      EST 01-AUG-2002
DEFINITION DKEZp686N12109_r1 763 (synonym: hlcc3) Homo sapiens cDNA clone
ACCESSION  AL698654
VERSION     AL698654
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 763)
AUTHORS    Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE      HRI human cDNA project
JOURNAL    Unpublished (2000)
COMMENT    Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5' & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
  1. .763
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="MAMMA1000851"
    /clone_lib="MAMMA1"
    /tissue_type="mammary gland"
    /note="Vector: pME185FL3"
BASE COUNT      137 a   205 c   260 g   158 t      3 others
ORIGIN

Query Match      25.3%; Score 89; DB 9; Length 763;
Best Local Similarity 100.0%; Pred. No. 2.1e-13;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 AATTCGAGCGAGTGTAGTGGCGCGGACCCGAGAGCGGAGCCCTTCTCTCC 323
|||||
DB 1 AATTCGAGCGAGTGTAGTGGCGCGGACCCGAGAGCGGAGCCCTTCTCTCC 60
|||||

RESULT 4
AL698654
LOCUS      AL698654      292 bp      mRNA      linear      EST 14-NOV-1994
DEFINITION HSC125081 normalized infant brain cDNA Homo sapiens cDNA clone
ACCESSION  Z44377
VERSION     Z44377
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 292)

```

AUTHORS Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Boulgatte, R., Jumelet, M.N., Lemay, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pieth, G., Pouliot, Y., Sebastiani-Kabakchis, C. and Tessier, A.

TITLE IMAGE: molecular integration of the analysis of the human genome and its expression

JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

MEDLINE 95277534

COMMENT Contact: Genethon
Genethon-Gerethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel.: 33169472800
Fax: 33160778698
Email: genethon@genethon.fr

Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: yic-12b08
Seq primer: (-21)K13.univser-sal.

FEATURES

Source

1..292
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-12b08"
/clone_lib="normalized infant brain cdna"
/sex="Female"
/tissue.type="total brain"
/dev_stage="3 months old"
/note="organ: Brain; Vector: lafmid BA; Site:1: HindIII;
Site:2: NotI; sex:Female; dev_stage=3 months old;
isolate:muscular atrophy patient; tissue.type=total brain
; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Souares, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"

BASE COUNT

ORIGIN 50 a 87 c 96 g 56 t
Query Match 21.3%; Score 75; DB 14; Length 292;
Best Local Similarity 100.0%; Pred. No. 9.6e-10;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 GTGAGTGGGCGGGACCGGAGCGAGCGCGACCGCTTCTCCCGGGTGCGGAGGG 337
|||||
Db 1 GTGAGTGGGCGGGACCGGAGCGAGCGCGACCGCTTCTCCCGGGTGCGGAGGG 60
|||||

QY 338 CAGGCGGGGAGCTC 352
|||||
Db 61 CAGGCGGGGAGCTC 75

RESULT 5

LOCUS BC034824
DEFINITION Homo sapiens, clone IMAGE:4749735, mRNA.
ACCESSION BC034824
VERSION BC034824.1 GI:21961568
KEYWORDS HTC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1004)
Strasbourg, R.
Direct Submission
Submitted (24-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.

cdna Library Preparation: Life Technologies, Inc.
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H.; Garcia, A.M.; Lu, X.; Hulyk, S.W.; Hale, S.M.;
Yoon, V.S.; Kowis, C.R.; Lawrence, S.; Martin, R.G.; Muzny, D.M.;
Richards, S.; Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 52 Row: d Column: 19
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5915657
This clone has the following problem: incomplete processing.

FEATURES

Source

1..1004
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4749735"
/tissue.type="Skin, squamous cell carcinoma"
/clone_lib="NCI CGAP_Skn4"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6.cdb"

BASE COUNT 306 a 234 c 230 g 234 t

Query Match 21.3%; Score 75; DB 11; Length 1004;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 75; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 278 GTGAGTGGGCGGGACCGGAGCGCGACCGCTTCTCCCGGGTGCGGAGGG 337
|||||
Db 1 GTGAGTGGGCGGGACCGGAGCGAGCGCGACCGCTTCTCCCGGGTGCGGAGGG 60
|||||

QY 338 CAGGCGGGGAGCTC 352
|||||
Db 61 CAGGCGGGGAGCTC 75

RESULT 6

LOCUS BC678861
DEFINITION mRNA sequence.
ACCESSION BC678861
VERSION BC678861.1 GI:13910258
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 998)
NIH-MGC <http://mgc.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cdna Library Preparation: Life Technologies, Inc.
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10603 Row: g Column: 16
High quality sequence stop: 860.
Location/Qualifiers
1..998
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES

Source

QY	31	CCCCACCCACCACTCCCCCACCTCCCTAGATCTGTGGTGGGGGCTGAACGTGGCC	90
		: :	
Db	905	SCCCCGGCCCGCCCGCCSCSSSSCCSSCCGCCGCCGCCCGCGSSGSSCGCGSSGSS	846
QY	91	CGTTAAGGGGGGGCGCGGCCGTCCTTCTGCTGAGTGACTGAACATACATAAAC	150
		: :	
Db	845	SGCGCCCGSGCGCCCGSGCGSCSSSCCGCGCGCGCGCGCGCGCGSGCGCGS	786
QY	151	AGAGCGCGGAAAGGGCGGGAGGAGGAGACACAGAGCTTACCGATAGTAACCTCT	210
		: :	
Db	785	SGCGCGSGCGCGSCCGCGSKCGCGCGSGGCGSSSGSGCGGGGSGCGSCSGCGSC	726
QY	211	CGCTCTGTCAGCGCAATCTATAAAGGAAGTAAGTCCCGGCAAAACCCGGAATTGCG	270
		: :	
Db	725	CGSSGSGCGSCGCCCGCGCGSCGCGSCGCGSGCGSGCGCGCGCGCGCGCGCGCC	666
QY	271	AGCGAGAGTAGTGGGCCCGGACCCGCACAGCGGACCGCAACCCTCTCTCCCGGGGTGC	330
		: :	
Db	565	SCSSGCGSSCGSCSSSCSSSCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG	606
QY	331	GGCAGGGCAGCGGGGAG	349
		: :	
Db	505	SCAGCGSSCMAGVAGSAG	587

RESULT 9	CNS0060N/c
LOCUS	
DEFINITION	
ACCESSION	CNS0060N
VERSION	910 bp
KEYWORDS	DNA
SOURCE	linear
ORGANISM	GSS 03-JUN-1999
REFERENCE	Drosophila melanogaster genome survey sequence 17 end of BAC #
AUTHORS	BACR14J21 of RPC1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
TITLE	AL065629
JOURNAL	AL065629.1 GI:4944598
COMMENT	GSS. Drosophila melanogaster. Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 910) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequences BP 191 91006 EVRY cedex - FRANCE (E-mail : seq-e@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osada and Aaron Manneser in Pictet de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPC1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strains Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila.bac.htm .

```

round at http://bacpat.med.buffalo.edu/drosophila/
location/Qualifiers
1. 910
/organism="Drosophila melanogaster"
/ab_xref="taxon:7227"
/clone="BACR14J21"
/clone_lib="RPCI-98"
/notes="end : T7"
BASE COUNT 202 a 63 c 112 g 198 t 335 others
ORIGIN

```

Query Match 12.0%; Score 42.4; DB 17; Length 910;
Best Local Similarity 13.8%; Pred. No. 0.53;

	Matches	44: Conservative	156: Mismatches	117: Indels	1: Gaps	1:
QY	34	CACCCACCACCTCCCCCAACTCCTAGATGTGCTGGCGGCCTAACGTCGCCCGT	93			
		: : : : : : : : : : : : : : : : : : : : :				
Dd	872	CACRSCVCSAMGAASAGAASAASSSAVSACGSSSSSVSVSSASASASAGYSRSASS	813			
QY	94	TTAAGGGCGGGCCCCGGCTCCAGTGCTTTCTGCTGAGTGACTGAACACTACATAAACA	153			
		: : : : : : : : : : : : : : : : : : : : : : :				
Dd	812	VSVSSVGSGSRKSASVSSSSSAGSAAVSAYAMACASMAVASAVMSASVVAARASAKS	753			
QY	154	GGCCGGGAACGGCGGGGAGGAGCACACAGGCTTTCACCGCATAGTAACCTC	213			
		: : : : : : : : : : : : : : : : : : : : :				
Dd	752	AASMARVAAVVARYASAAVVASAAVAAVAASVASASASVMASSSSSSSSSSSS	693			
QY	214	CTCGGTCGACCGGAATCTATAA-AAGGAACTAGTCGCCGGCAAACACCCGTAATTGGC	272			
		: : : : : : : : : : : : : : : : : : : : : : : :				
Dd	692	SASSSASASGMVSCRNAVSVASASAAASARSASRASAVMAVAASASSSASVSSGS	633			
QY	273	CGAGAGTGTGGGGCGGGACCCCGAGCGAGCGACCTCTCTCCCGGCTCGG	332			
		: : : : : : : : : : : : : : : : : : : : :				
Dd	632	SCSSRRCCSASSSSASASGSSSSSSSSSSSSSCGGSCASCASSSSSRSGCGCSSVC	573			
QY	333	CAGGCACGGCGGGAGC	350			
		: : : : : : : : : : : : : : : : : : : : : : : :				
Dd	572	SSCGSSSSSSSSSSSSSGC	555			

RESULT 10
CNS0091P
LOCUS
CNS0091P DNA linear GSS 03-JUN-1999
925 bp
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of PAC #
BACR9D16 of RCI-38 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
ALC53013
VERSION
ALC53013.1 GI:4934461
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster.
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 925)
REFERENCE

REFERENCE	unpublished, Drosophilinae, Drosophilina.
AUTHORS	1 (bases 1 to 925)
TITLE	Genoscope.
DATE	Direct Submission
JOURNAL	Submitted (02-JUN-1999): Genoscope - Centre National de Sequençage :
	BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr
	- Web : www.genoscope.cns.fr)
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named pC1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp. The same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .

FEATURES	SOURCE
round at	locatp//qualifiers
round at	i. . 925
round at	/organism="Drosophila melanogaster"
round at	/db_xref="taxon:7227"
round at	/close="BACR19D1c"
round at	/close_lib="RPI-98"
round at	/note="end : TE13"
BASE COUNT	120 a 61 c 61 g 172 t 511 others
ORIGIN	

Query Match
12.08; score 42.2; DB 17; Length 925;

```

Best Local Similarity 12.5%; Pred. No. 0.6;
Matches 4; Conservative 148; Mismatches 136; Indels 0; Gaps 0;

Qy 23 CCCTCCACCCACCCACACCTCCCCCACTCCCTAGATGTCTCTGGGGGCTGA 82
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 CCGSSSSCCGCCCCCSCSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 637
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 83 ACCTCGCCCTTAAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 638 SSSSSSTSSSTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 697
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 143 ACATAAACAGAGCGCGGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 698 SSSSSSTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 757
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 203 TAACCTCTGGCTCGGTCCACCCCAATCTA-AAAAGAACTAGTCCCGGAAACCCG 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 758 CCGCTCCCSYSSSTSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 817
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 263 TAATTCGAGCGAGTCAAGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 818 YEMCYTSTCGGGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 877
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 323 CCGCTCGCGGAGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 349
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 878 SSKSSASSSSSVSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 904
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
CNS006N3/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BAC14H04 of RPCI-98 library from Drosophila melanogaster (fruit
fly); genomic survey sequence.
ACCESSION
AL065775
VERSION
AL065775.1 GI:4944655
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster.
ORGANISM
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 912)
AUTHORS
Genoscope.
DIRECT SUBMISSION
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : secref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
1..912
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BAC14H04"
/clone.lib="RPCI-98"
/note="end : TE13"

BASE COUNT
179 a 240 c 82 g 146 t 266 others
ORIGIN
Qy 23 CCCTCCACCCACCCACACCTCCCCCACTCCCTAGATGTCTCTGGGGGCTGA 82
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 817 CYCCCCSCCCCCSCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 876

```


<hr/>					
RESULT 5					
AU121731		763 bp	mRNA	linear	EST 01-AUG-2002
LOCUS					
DEFINITION	AU121731 MAMMAL Homo sapiens cDNA clone MAMMA1000851 5', mRNA sequence.				
ACCESSION	AU121731				
VERSION	AU121731.1	GI:10936966			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 763)				
AUTHORS	Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and Isogai,T.				
TITLE	HRI human cDNA project				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Takao Isoqai Genomics Laboratory Helix Research Institute 1532-3 Yara, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3975 Fax: 81-438-52-3986 Email: genomics@hri.co.jp HRI human cDNA project: 5'- s 3'-end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute. Location/Qualifiers 1..763 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="MAMMA1000851" /clone_lib="MAMMA1" /tissue_type="mammary gland" /note="Vector: pME18SP13"				
FEATURES					
source	137 a	205 c	260 g	158 t	3 others
BASE COUNT					
ORIGIN					
Query Match	5.4%	Score 89;	DB 9;	Length 763;	
Best Local Similarity	100.0%;	Pred. No. 7.5e-08;			
Matches	89;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1555	AATTGGAGCGAGAGTGATGTCGGCCGGACCGCAGAGCGAGCGACCTTCTCTCC	1614		
Dd	1	AATTGGAGCGAGAGTGATGTCGGCCGGACCGCAGAGCGAGCGACCTTCTCTCC	60		
QY	1615	GGGTGCGCGAGCGGCAGGGCGGGAGCTC	1643		
Dd	61	GGGTGCGCGAGCGGCAGGGCGGGAGCTC	89		
RESULI	6				
LOCUS	244377				
DEFINITION	HSC1ZB081 normalized infant brain cDNA Homo sapiens cDNA clone c-1zb08, mRNA sequence.				
ACCESSION	244377				
VERSION	244377.1	GI:573506			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 292)				
AUTHORS	Aufrey,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes,M.D., Duprat,S., Houllgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F., Mitchell,E., Marriage-Samsom,R., Pietu,G., Pouliot,Y., Sebastiani-Kabatchis,C. and Tessier.A.				
TITLE	IMAGE: molecular integration of the analysis of the human genome				

TITLE The zinc finger protein 202 (ZNF202) is a transcriptional repressor of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene expression and a modulator of cellular lipid efflux

JOURNAL J. Biol. Chem. 276 (15), 12427-12433 (2001)

MEDLINE 21192304

PUBMED 11279031

REFERENCE 2 (bases 1 to 1167)

AUTHORS Porsch-Oezgueruermez, M.K.

TITLE Direct Submission

JOURNAL Submitted (05-JAN-2000) Porsch-Oezgueruermez M.K., Institute for Clinical Chemistry, University of Regensburg, Franz-Josef-Strauss-Allee 11, 93042 Regensburg, GERMANY

FEATURES

source Location/Qualifiers

1..1167

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/map="9422-31"

/cell_type="leukocyte"

gene 1..1167

/gene="ABC-1"

promoter 1..942

/gene="ABC-1"

/function="cholesterol efflux regulatory protein"

TATA_signal 896..900

/gene="ABC-1"

intron 1148..1167

/gene="ABC-1"

/function="cholesterol efflux regulatory protein"

BASE COUNT 278 a 315 c 327 g 247 t

ORIGIN

Query Match 100.0%; Score 139; DB 9; Length 1167;

Best Local Similarity 100.0%; Pred. No. 1.2e-31;

Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCCCCCGCTCCACGTGCTTCTGCTGAGTGAAGTAACTACATAAAGAGCCCGGAA 60

|||||

Db 768 GGCCCCCGCTCCACGTGCTTCTGCTGAGTGAAGTAACTACATAAAGAGCCCGGAA 827

QY 61 CGGGCGGGGAGGAGGAGACAGAGCTTGACCATAGTAACTACATAAAGAGCCCGGAA 120

|||||

Db 828 CGGGCGGGGAGGAGGAGACAGAGCTTGACCATAGTAACTACATAAAGAGCCCGGAA 120

QY 121 GCCGAATCTATAAAGGAA 139

|||||

Db 888 GCCGAATCTATAAAGGAA 906

RESULT 2

AX060715

LOCUS AX060715 1643 bp DNA linear PAT 22-JAN-2001

DEFINITION Sequence 3 from Patent WO0078972.

ACCESSION AX060715

VERSION AX060715.1 GI:12406104

KEYWORDS human.

SOURCE

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1643)

Law, R.M., Wade, D., and Garvin, M.

Regulation with binding cassette transporter protein abcl

Patent: WO 0078972-A 3 28-DEC-2000;

CV THERAPEUTICS, INC. (US)

FEATURES

source Location/Qualifiers

1..1643

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 370 a 413 c 457 g 403 t

ORIGIN

Query Match 100.0%; Score 139; DB 6; Length 1643;

Best Local Similarity 100.0%; Pred. No. 1.1e-31;

Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCCCCCGCTCCACGTGCTTCTGCTGAGTGAAGTAACTACATAAAGAGCCCGGAA 60

|||||

Db 1394 GGCCCCCGCTCCACGTGCTTCTGCTGAGTGAAGTAACTACATAAAGAGCCCGGAA 1453

QY 61 CGGGCGGGGAGGAGGAGACAGAGCTTGACCATAGTAACTACATAAAGAGCCCGGAA 120

|||||

Db 1454 CGGGCGGGGAGGAGGAGACAGAGCTTGACCATAGTAACTACATAAAGAGCCCGGAA 1513

QY 121 GCCGAATCTATAAAGGAA 139

|||||

Db 1514 GCCGAATCTATAAAGGAA 1532

RESULT 4

AC021246

LOCUS AC021246 69570 bp DNA linear HTG 13-JUL-2000

DEFINITION Homo sapiens clone RP11-IN10, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC021246

VERSION AC021246.2 GI:9119882

KEYWORDS HTG; HTGS_PHASE0.

SOURCE

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 69570)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome, clone RP11-IN10

Unpublished

REFERENCE 2 (bases 1 to 69570)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Best Local Similarity 100.0%; Pred. No. 1.1e-31;

Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCCCCCGCTCCACGTGCTTCTGCTGAGTGAAGTAACTACATAAAGAGCCCGGAA 60

|||||

Db 1394 GGCCCCCGCTCCACGTGCTTCTGCTGAGTGAAGTAACTACATAAAGAGCCCGGAA 1453

QY 61 CGGGCGGGGAGGAGGAGACAGAGCTTGACCATAGTAACTACATAAAGAGCCCGGAA 120

|||||

Db 1454 CGGGCGGGGAGGAGGAGACAGAGCTTGACCATAGTAACTACATAAAGAGCCCGGAA 1513

QY 121 GCCGAATCTATAAAGGAA 139

|||||

Db 1514 GCCGAATCTATAAAGGAA 1532

RESULT 3

AX060894

LOCUS AX060894 1643 bp DNA linear PAT 22-JAN-2001

DEFINITION Sequence 3 from Patent WO0078971.

ACCESSION AX060894

VERSION AX060894.1 GI:12406271

KEYWORDS human.

SOURCE

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1643)

Law, R.M., Wade, D., Oram, J.F. and Garvin, M.

ATP binding cassette transporter protein abcl polypeptides

Patent: WO 0078972-A 3 28-DEC-2000;

CV THERAPEUTICS, INC. (US)

FEATURES

source Location/Qualifiers

1..1643

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 370 a 413 c 457 g 403 t

ORIGIN

Query Match 100.0%; Score 139; DB 6; Length 1643;

Best Local Similarity 100.0%; Pred. No. 1.1e-31;

Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCCCCCGCTCCACGTGCTTCTGCTGAGTGAAGTAACTACATAAAGAGCCCGGAA 60

|||||

Db 1394 GGCCCCCGCTCCACGTGCTTCTGCTGAGTGAAGTAACTACATAAAGAGCCCGGAA 1453

QY 61 CGGGCGGGGAGGAGGAGACAGAGCTTGACCATAGTAACTACATAAAGAGCCCGGAA 120

|||||

Db 1454 CGGGCGGGGAGGAGGAGACAGAGCTTGACCATAGTAACTACATAAAGAGCCCGGAA 1513

QY 121 GCCGAATCTATAAAGGAA 139

|||||

Db 1514 GCCGAATCTATAAAGGAA 1532

RESULT 4

AC021246

LOCUS AC021246 69570 bp DNA linear HTG 13-JUL-2000

DEFINITION Homo sapiens clone RP11-IN10, LOW-PASS SEQUENCE SAMPLING.

ACCESSION AC021246

VERSION AC021246.2 GI:9119882

KEYWORDS HTG; HTGS_PHASE0.

SOURCE

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 69570)

Birren, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome, clone RP11-IN10

Unpublished

REFERENCE 2 (bases 1 to 69570)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,
Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeAtellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Hesford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lepocky, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Marcuis, N., McEvan, P., McGurk, A., McKernan, K.,
McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.

TITLE
JOURNAL

COMMENT

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gl:6705871.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seg.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2512
Center clone name: L_N10

NOTE: This record contains 73 individual
sequencing reads that have not been assembled into
contigs. Runs of N are used to separate the reads
and the order in which they appear is completely
arbitrary. Low-pass sequence sampling is useful for
identifying clones that may be gene-rich and allows
overlap relationships among clones to be deduced.
However, it should not be assumed that this clone
will be sequenced to completion. In the event that
the record is updated, the accession number will
be preserved.

1 871: contig of 571 bp in length
872 971: gap of 100 bp
972 1834: contig of 863 bp in length
1835 1934: gap of 100 bp
1935 2804: contig of 870 bp in length
2805 2904: gap of 100 bp
2905 3745: contig of 841 bp in length
3746 3845: gap of 100 bp
3846 4596: contig of 851 bp in length
4597 4796: gap of 100 bp
4797 5640: contig of 844 bp in length
5641 5740: gap of 100 bp
5741 6540: contig of 800 bp in length
6541 6640: gap of 100 bp
6641 7509: contig of 869 bp in length
7510 7609: gap of 100 bp
7610 8473: contig of 870 bp in length
8480 8579: gap of 100 bp
8580 9430: contig of 851 bp in length
9431 9530: gap of 100 bp
9531 10376: contig of 846 bp in length
10377 10476: gap of 100 bp
10477 11322: contig of 846 bp in length
11323 11422: gap of 100 bp
11423 12303: contig of 880 bp in length
12303 12402: gap of 100 bp
12403 13280: contig of 878 bp in length
13281 13380: gap of 100 bp
13381 14241: contig of 861 bp in length
14242 14341: gap of 100 bp

14342 15197 15 855 bp in length
15197 16123: contig of 827 bp in length
16123 16223: gap of 100 bp
16224 17072: contig of 849 bp in length
17073 17172: gap of 100 bp
17173 18041: contig of 869 bp in length
18042 18141: gap of 100 bp
18142 19009: contig of 868 bp in length
19010 19109: gap of 100 bp
19110 19966: contig of 857 bp in length
19967 20056: gap of 100 bp
20057 20921: contig of 855 bp in length
20922 21021: gap of 100 bp
21022 21865: contig of 844 bp in length
21866 21955: gap of 100 bp
21956 22832: contig of 867 bp in length
22833 22932: gap of 100 bp
22933 23780: contig of 848 bp in length
23781 23880: gap of 100 bp
23881 24733: contig of 853 bp in length
24734 24833: gap of 100 bp
24834 25670: contig of 837 bp in length
25671 25770: gap of 100 bp
25771 26621: contig of 851 bp in length
26622 26721: gap of 100 bp
26722 27576: contig of 855 bp in length
27577 27676: gap of 100 bp
27677 28532: contig of 856 bp in length
28533 28632: gap of 100 bp
28633 29492: contig of 860 bp in length
29493 29592: gap of 100 bp
29593 30455: contig of 863 bp in length
30456 30555: gap of 100 bp
30556 31410: contig of 855 bp in length
31411 31510: gap of 100 bp
31511 32368: contig of 858 bp in length
32369 32468: gap of 100 bp
32469 33312: contig of 844 bp in length
33313 33412: gap of 100 bp
33413 34268: contig of 856 bp in length
34269 34368: gap of 100 bp
34369 35204: contig of 836 bp in length
35205 35304: gap of 100 bp
35305 36156: contig of 852 bp in length
36157 36256: gap of 100 bp
36257 37128: contig of 872 bp in length
37129 37228: gap of 100 bp
37229 38083: contig of 855 bp in length
38084 38183: gap of 100 bp
38184 39031: contig of 848 bp in length
39032 39131: gap of 100 bp
39132 40006: contig of 875 bp in length
40007 40106: gap of 100 bp
40107 40967: contig of 861 bp in length
40968 41067: gap of 100 bp
41068 41913: contig of 846 bp in length
41914 42013: gap of 100 bp
42014 42824: contig of 811 bp in length
42825 42924: gap of 100 bp
42925 43776: contig of 852 bp in length
43777 43876: gap of 100 bp
43877 44752: contig of 876 bp in length
44753 44852: gap of 100 bp
44853 45724: contig of 872 bp in length
45725 45824: gap of 100 bp
45825 46643: contig of 819 bp in length
46644 46743: gap of 100 bp
46744 47599: contig of 856 bp in length
47599 47699: gap of 100 bp
47700 48551: contig of 852 bp in length
48552 48651: gap of 100 bp
48652 49485: contig of 834 bp in length

```

* 49486 49585: gap of 100 bp
* 49586 50440: contig of 855 bp in length
* 50441 50540: gap of 100 bp
* 50541 51404: contig of 864 bp in length
* 51405 51504: gap of 100 bp
* 51505 52377: contig of 868 bp in length
* 52378 52477: gap of 100 bp
* 52478 53328: contig of 856 bp in length
* 53329 53428: gap of 100 bp
* 53429 54258: contig of 840 bp in length
* 54259 54368: gap of 100 bp
* 54369 55223: contig of 861 bp in length
* 55224 55323: gap of 100 bp
* 55324 56197: contig of 868 bp in length
* 56198 56297: gap of 100 bp
* 56298 57163: contig of 866 bp in length
* 57164 57263: gap of 100 bp
* 57264 58130: contig of 867 bp in length
* 58131 58230: gap of 100 bp
* 58231 59082: contig of 852 bp in length
* 59083 59182: gap of 100 bp
* 59183 60020: contig of 838 bp in length
* 60021 60120: gap of 100 bp
* 60121 60983: contig of 853 bp in length
* 60984 61083: gap of 100 bp
* 61084 61935: contig of 852 bp in length
* 61936 62035: gap of 100 bp
* 62036 62866: contig of 831 bp in length
* 62867 62966: gap of 100 bp
* 62967 63827: contig of 861 bp in length
* 63828 63927: gap of 100 bp
* 63928 64783: contig of 856 bp in length
* 64784 64883: gap of 100 bp
* 64884 65740: contig of 857 bp in length
* 65741 65840: gap of 100 bp
* 65841 66684: contig of 844 bp in length
* 66685 66784: gap of 100 bp
* 66785 67651: contig of 867 bp in length
* 67652 67751: gap of 100 bp

Query Match 100.0%; Score 139; DB 2; Length 59570;
Best Local Similarity 100.0%; Pred. No. 7.6e-32;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 1 GGGCCCCGGCTCCAGTCTTCTCTGAGTGACTGAACATACATAACAGAGGCGCGGAA 60
|||||
Db 41406 GGGCCCCGGCTCCAGTCTTCTCTGAGTGACTGAACATACATAACAGAGGCGCGGAA 41465

QY 61 CGGGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
|||||
Db 41466 CGGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 41525

QY 121 GCCGATCTATAAAGGAA 139
|||||
Db 41526 GCCGATCTATAAAGGAA 41544

RESULT 5
LOCUS AL359182/c 96717 bp DNA linear PRI 11-JAN-2002
DEFINITION Human DNA sequence from clone Rp11-217B7 on chromosome 9, complete
sequence.
ACCESSION AL359182
VERSION AL359182.20 GI:18151453
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 96717)
Direct Submission
AUTHORS Skuce,C.
TITLE Submitted (11-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL

```

COMMENT

Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jan 15, 2002 this sequence version replaced g1:18121468. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TrEMBL; Wp., WormPEP; information on the WormPEP database can be found at <http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/Ref/Chr9> Rp11-217B7 is from the library RpC1-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAGE3.6

IMPORTANT: This sequence is not the entire insert of clone Rp11-217B7 it may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone Rp11-217B7 is at 96717 in this sequence. The true left end of clone Rp11-122F10 is at 72980 in this sequence. The true right end of clone Rp11-31J20 is at 2000 in this sequence.

FEATURES

```

source Location/Qualifiers
1..96717
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="Rp11-217B7"
/clone_lib="RPC1-11.1"
3238..3278
misc_feature
84249..84273
/note="Sequence from overlapping clone Rp11-122F10 (AC026643). Assembly confirmed by restriction digest."
92050..92163
misc_feature
/note="Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
92411..92557
misc_feature
/note="Sequence from reads from a short insert library derived from a single pUC clone. Restriction digest data confirm the assembly."
BASE COUNT 27673 a 21138 c 20380 g 27526 t
ORIGIN
Query Match 100.0%; Score 139; DB 9; Length 96717;
Best Local Similarity 100.0%; Pred. No. 7.3e-32;
Matches 139; Conservative 0; Mismatches 0; Indels 0; Gaps 0:
QY 1 GGGCCCCGGCTCCAGTCTTCTCTGAGTGACTGAACATACATAACAGAGGCGCGGAA 60
|||||
Db 17416 GGGCCCCGGCTCCAGTCTTCTCTGAGTGACTGAACATACATAACAGAGGCGCGGAA 17357
QY 61 CGGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
|||||
Db 17356 CGGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 17297

```

```

Oy 121 GCCGAATCTATTAAGGAA 139
D5 17296 GCCGAATCTATTAAGGAA 17278

RESULT 6
AC012230
LOCUS
DEFINITION
AC012230 175064 bp DNA linear HTG 22-APR-2000
Homo sapiens clone RP11-1M10, WORKING DRAFT SEQUENCE, 39 unordered
pieces.
AC012230
VERSION
AC012230.3 GI:7637254
KEYWORDS
HTG, HTGS-PHASE1, HTGS-DRAFT.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 175064)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-1M10
Unpublished
2 (bases 1 to 175064)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balwin,J., Barna,N., Beckerly,R., Bozulavkiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donegan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funk,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heath,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Melidrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 22, 2000 this sequence version replaced gi:6454033.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2510
Center clone name: L_M_10
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 117571 bases at least Q40
Consensus quality: 145745 bases at least Q30
Consensus quality: 160940 bases at least Q20
Insert size: 185000; agarose-ftp
Insert size: 171264; sum-of-contigs
Quality coverage: 2.9 in Q20 bases; agarose-ftp
Quality coverage: 3.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 1004 1103: gap of 1003 bp in length
* 1004 1103: gap of 100 bp
* 145592 157391: contig of 11800 bp in length

```



```

Db 2838 CCTCTGGCTGGTGGAGCCGCAATCTATAAAGCACTAGTCCCGCAAAAACCCCGCTAA 2897
QY 1557 TTCCGAGGAGAGTGAAGTGGGGCGGAGACCCGAGAGCCGAGCCGACCCCTCTCTCCCG 1616
|||||
Db 2898 TTCCGAGGAGAGTGAAGTGGGGCGGAGACCCGAGAGCCGAGCCGACCCCTCTCTCCCG 2957
|||||
QY 1617 GCTGCGGCGAGGCGAGCGCGGAGGAGTC 1643
|||||
Db 2958 GCTGCGGCGAGGCGAGCGCGGAGGAGTC 2964
|||||

RESULT 6
AF275948
LOCUS AF275948 149034 bp DNA linear PRI 17-JUL-2000
DEFINITION Homo sapiens ABCA1 (ABCA1) gene, complete cds.
ACCESSION AF275948
VERSION AF275948.1 GI:9247085
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 149034)
AUTHORS Santamarina-Fojo,S., Peterson,K., Knapper,C., Olu,X., Freeman,L.,
Cheng,J.F., Ogrio,J., Remaley,A., Yang,X.P., Haudenschild,G.C.,
Prades,C., Chamin,L.G., Blackmon,E., Francois,T., Duverger,N.,
Rubin,E.M., Rosier,M., Denerlie,P., Fredrickson,D.S. and Brewer,H.B.
Jr.
Complete genomic sequence of the human ABCA1 gene: analysis of the
human and mouse Abp-binding cassette A promoter
Proc. Natl. Acad. Sci. U.S.A. 97 (14), 7987-7992 (2000)
20345099
10884428
TITLE 2 (bases 1 to 149034)
AUTHORS Santamarina-Fojo,S., Peterson,K.M., Knapper,C.L., Freeman,L.A.,
Remaley,A.T., Yang,X.P., Haudenschild,G.C., Blackmon,E.B.,
Francois,T.L. and Brewer,H.B. Jr.
Direct Submission
Submitted (08-JUN-2000) Molecular Disease Branch, National
Institutes of Health, National Heart, Lung and Blood Institute,
Bethesda, MD 20892, USA
FEATURES
Location/Qualifiers
1..149034
/organism="Homo sapiens"
/cb_xref="taxon:9606"
1454..148034
/gene="ABCA1"
join(1454..1674,25831..25999,40385..40478,45012..45153,46423..46541,
46423..46541,67729..67850,70831..71007,83960..84052,89010..89250,
89010..89250,91962..92101,92433..92349,96758..96955,
97702..97907,98428..98604,100391..100613,102360..102581,
103642..103846,104951..105064,106862..107033,
108023..108154,109460..109502,109806..109943,
110646..110866,112124..112196,113183..113385,
115053..115101,115297..115410,116806..116954,
118604..118728,119964..120062,123094..123283,
124804..124898,126208..126240,127363..127469,
128943..129017,129339..129708,130939..131116,
131114..133229,133341..133485,134526..134649,
135736..135865,136130..136250,137038..137100,
138008..138114,138469..140610,140982..141116,
142080..142163,142646..142738,143397..143640,
144581..145034)
/gene="ABCA1"
/product="ABCA1"
join(1454..1674,25831..25923)
/gene="ABCA1"
repeat_region 2406..2678
repeat_region 4240..4509
repeat_region 4998..5270
/pt_family="Alu"

```


GenCore version 5.1.3
Copyright: (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 14:33:17 : Search time 92.7366 Seconds
(without alignments)
7959.555 Million cell updates/sec

Title: US-09-596-141c-3

Perfect score: 1643

Sequence: 1 gaattccttgccgggtggtc.....caggcgaggcggggagatc 1643

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 396772 seqs, 224532407 residues

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA:*

```
1: /cgn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2.6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2.6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
4: /cgn2.6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2.6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2.6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2.6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2.6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2.6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2.6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2.6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2.6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2.6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2.6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1485.4	90.4	3231	10	US-09-846-456-1 Sequence 1, Appli
2	1394.4	84.9	2893	10	US-09-846-456-3 Sequence 3, Appli
3	102.2	6.2	2893	10	US-09-846-456-3 Sequence 3, Appli
4	102.2	6.2	3231	10	US-09-846-456-1 Sequence 1, Appli
5	91	5.5	221	10	US-09-846-456-4 Sequence 4, Appli
6	43.4	2.6	32190	9	US-09-850-670-255 Sequence 255, App
7	43.4	2.6	32249	9	US-09-860-670-260 Sequence 260, App
8	43	2.6	366	9	US-09-796-692-8332 Sequence 8332, Ap
9	41.8	2.5	251	10	US-09-764-855-6C Sequence 60, Appl
10	41.6	2.5	1974	10	US-09-834-975-994 Sequence 994, App
11	41.4	2.5	2099	10	US-09-764-870-14 Sequence 14, Appl
12	41	2.5	12718	10	US-09-764-877-3972 Sequence 3972, Ap
13	40.6	2.5	545	10	US-09-878-574-4299 Sequence 4299, Ap
14	40.2	2.4	1372	9	US-10-008-118A-19 Sequence 19, Appl
15	40.2	2.4	1372	10	US-09-443-704-19 Sequence 19, Appl
16	40	2.4	16511	10	US-09-764-869-2064 Sequence 2064, Ap
17	39.6	2.4	2756	10	US-09-925-30-351 Sequence 351, App
18	39.4	2.4	335	10	US-09-834-975-453 Sequence 453, App
19	39	2.4	668	10	US-09-731-872-192 Sequence 192, App

```

C 20 39 2.6 2277 10 US-09-917-800A-1513 Sequence 1513, Ap
C 21 38.6 2.3 1338 10 US-09-800-723-44 Sequence 44, Appl
C 22 38.4 2.3 295 10 US-09-764-846-33 Sequence 33, Appl
C 23 38.4 2.3 304 10 US-09-764-846-109 Sequence 109, App
C 24 38.4 2.3 310 10 US-09-864-864-162 Sequence 162, App
C 25 38.4 2.3 2504 10 US-09-729-674-147 Sequence 147, App
C 26 38.4 2.3 32195 10 US-09-764-870-611 Sequence 611, App
C 27 38.4 2.3 32195 10 US-09-764-870-617 Sequence 617, App
C 28 38.4 2.3 32195 10 US-09-764-869-1605 Sequence 1605, App
C 29 38.2 2.3 2509 10 US-09-925-301-540 Sequence 540, App
C 30 38 2.3 1113 10 US-09-834-975-739 Sequence 739, App
C 31 38 2.3 1484 9 US-09-932-598-292 Sequence 292, App
C 32 38 2.3 1484 9 US-09-989-293A-292 Sequence 292, App
C 33 38 2.3 1484 9 US-09-989-735-292 Sequence 292, App
C 34 38 2.3 1484 9 US-09-990-444-292 Sequence 292, App
C 35 38 2.3 1484 9 US-09-989-730-292 Sequence 292, App
C 36 38 2.3 1484 9 US-09-990-436-292 Sequence 292, App
C 37 38 2.3 1484 9 US-09-991-181-292 Sequence 292, App
C 38 38 2.3 1484 9 US-09-993-687-292 Sequence 292, App
C 39 38 2.3 1484 9 US-09-989-734-292 Sequence 292, App
C 40 38 2.3 1484 9 US-09-997-653-292 Sequence 292, App
C 41 38 2.3 1484 9 US-10-774-590-219 Sequence 219, App
C 42 38 2.3 1484 9 US-10-776-758-219 Sequence 219, App
C 43 38 2.3 1484 9 US-10-175-737-219 Sequence 219, App
C 44 38 2.3 1484 10 US-09-989-722-292 Sequence 292, App
C 45 38 2.3 1484 10 US-09-989-723-292 Sequence 292, App

```

ALIGNMENTS

RESULT 1

US-09-846-456-1

; Sequence 1, Application US/09846456

; Patent No. US20020146792A1

; GENERAL INFORMATION:

; APPLICANT: Rosier, Marie

; APPLICANT: Prades, Catherine

; APPLICANT: Lemoine, Cendrine

; APPLICANT: Naudin, Laurent

; APPLICANT: Benefle, Patrice

; APPLICANT: Duverger, Nicolas

; APPLICANT: Brewer, Bryan

; APPLICANT: Remaley, Alan

; APPLICANT: Fojo, Silvia

; TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying

; TITLE OF INVENTION: Activity and Therapeutic Uses

; FILE REFERENCE: 3806.0505

; CURRENT APPLICATION NUMBER: US/09/846,456

; CURRENT FILING DATE: 2001-05-02

; PRIOR APPLICATION NUMBER: US 60/201,280

; PRIOR FILING DATE: 2000-05-02

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 1

; LENGTH: 3231

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-846-456-1

Query Match 50.4%; Score 1485.4; DB 10; Length 3231;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 1615; Conservative 0; Mismatches 6; Indels 26; Gaps 10;

QY 16 GGCTCCACATGCCACTTCCAGGGCGCTGCTTGG--CCTCTTCTATGGGCTGCTCTGAGTGT 73

Db 1345 GCGTCCACATGCCACTTCCAGGGCGCTGCTTGGGCTCTCTCTATGCTGCTCTGAGTGT 1404

QY 74 GATAGAACCACTGATGTGAGTACCTGGCTTGAG-CGFGGCGCTGGAGATCCTGTTGACTG 132

Db 1405 GATAGAACCACTGATGTGAGTACCTGGGCTTGAGCCCTGGCGGAGATCCTGTTGACTG 1464

QY 133 TAGCATGAGAGGGGCTTTGT-CAGCTGAATGTTCTGATGCAGGTGGTGGGAGTTCGTGAAT 191

```
Db 1465 TAGCATGGAGGGGCTGTGCGAGCTGAATGTCTGCATGAGGTGGAGTCTGCAAT 1524
QY 192 ATGATGGAGCTGGAGTGGGAGAGAGTAGGCTTGGGCGAGCTCTCTCATGCACTCA 251
Db 1525 ATGATGGAGCTGGAGTGGGAGAGAGTAGGCTTGGGCGAGCTCTCTCATGCACTCA 1584
QY 252 TTCGCGCCAAACTCAGCTCAACTGTGAGAGCTGAATGTGAATCTGCCCTTCAAGGT 311
Db 1585 TTCGCGCCAAACTCAGCTCAACTGTGAGAGCTGAATGTGAATCTGCCCTTCAAGGT 1644
QY 312 GGTCAAAAGGTATCTTTGTCAGGTAGAGAGCTTGTGCGCTCCAGCTGCACCTCCAGG 371
Db 1645 GGTCAAAAGGTATCTTTGTCAGGTAGAGAGCTTGTGCGCTCCAGCTGCACCTCCAGG 1704
QY 372 GCGTGCTTGGGCTCTCTACGGGTGTGCTGAGCTTCTATGATCTCCCTTCAGGGC 430
Db 1705 GCGTGCTTGGGCTCTCTACGGGTGTGCTGAGCTTCTATGATCTCCCTTCAGGGC 1761
QY 431 AGATTCATATTTAGACTCTTCACAGTTTGCACCTGAGTTTGGCCAGAAATAAGTGCAT 490
Db 1762 AGATTCATATTTAGACTCTTCACAGTTTGCACCTGAGTTTGGCCAGAAATAAGTGCAT 1821
QY 491 TAGTTTGTGGCTTGATGATGACATAATATTAGACATATGTTGTTAGGCTTCAT 550
Db 1822 TAGT-TGTTGCGCTTGATGATGACATAATATTAGAC--ATGGTGTAGGCTTCAT 1879
QY 551 CTAATCTTGGCTTTT-TTTCGCGCTCCAGTGTGTTGGGTATGTTTGGCTTCATACAG 610
Db 1880 CTAATCTTGGCTTTT-TTTCGCGCTCCAGTGTGTTGGGTATGTTTGGCTTCATACAG 1938
QY 611 CCAAGGCAACAGATAAGCTTCAGGCTCGAGTGGCTGCTACATAATTTACAGCACTGCAA 670
Db 1939 CCAAGGCAACAGATAAGCTTCAGGCTCGAGTGGCTGCTACATAATTTACAGCACTGCAA 1998
QY 671 TTCTCTGGCTGCACCTTCACAAATGATACAACTAAATCAAGTCTGTGTTTATCAC 730
Db 1999 TTCTCTGGCTGCACCTTCACAAATGATACAACTAAATCAAGTCTGTGTTTATCAC 2058
QY 731 AGGAGGCTGATCAATAATGAATTAATAAGGGGCTGGTCCCAATATGTTCTGTGTT 790
Db 2059 AGGAGGCTGATCAATAATGAATTAATAAGGGGCTGGTCCCAATATGTTCTGTGTT 2117
QY 791 TTG-----TTGTTGTTGTTCTTTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 841
Db 2118 TTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2177
QY 842 TGAAGAGAGAGCTAAGATGTTCCCTCGGGTCCCTGTAGGGAGCTGGGAGCTCAGGCT 901
Db 2178 TGAAGAGAGAGCTAAGATGTTCCCTCGGGTCCCTGTAGGGAGCTGGGAGCTCAGGCT 2237
QY 902 GGAATCTCCAGGCGAGTAGTGGCTATCAAAATCAAAATCAAGTCTGAGTTTGGGGGAA 961
Db 2238 GGAATCTCCAGGCGAGTAGTGGCTATCAAAATCAAAATCAAGTCTGAGTTTGGGGGAA 2297
QY 962 AACAAAGCAGCCCAATACCCAGAGAGCTGTCCGGCTTCCCTCAGCTCCAGCTAGGCT 1021
Db 2298 AACAAAGCAGCCCAATACCCAGAGAGCTGTCCGGCTTCCCTCAGCTCCAGCTAGGCT 2357
QY 1022 TTCAAGGAAACAAAGAGAGAGCAAAATGATGCGCTCCTCAGGAGATTTCAGCTAGA 1081
Db 2358 TTCAAGGAAACAAAGAGAGAGCAAAATGATGCGCTCCTCAGGAGATTTCAGCTAGA 2417
QY 1082 GCTCTCTCTCCCAATCCCTCCCTCGGCTGAGGAACTACAAAGGAGGAGGAGGAGG 1141
Db 2418 GCTCTCTCTCCCAATCCCTCCCTCGGCTGAGGAACTACAAAGGAGGAGGAGGAGG 2477
QY 1142 CGGAAGCAGGATTAGAGAGAGCAAAATGATGCGCTCCTCAGGAGATTTCAGCTAGA 1201
Db 2478 CGGAAGCAGGATTAGAGAGAGCAAAATGATGCGCTCCTCAGGAGATTTCAGCTAGA 2537
QY 1202 ACTAGAGATCTGCGGCGAGCCCGAGCCAGCTTCCCGGCGGTCTTACGGCGCGG 1261
```

RESULT 2

```
US-09-846-456-3
; Sequence 3, Application US/09846456
; Patent No. US20020146792A1
; GENERAL INFORMATION:
; APPLICANT: Rosier, Marie
; APPLICANT: Prades, Catherine
; APPLICANT: Lemoine, Catherine
; APPLICANT: Naudin, Laurent
; APPLICANT: Deneffe, Patrice
; APPLICANT: Duverger, Nicolas
; APPLICANT: Brewer, Bryan
; APPLICANT: Remaley, Alan
; APPLICANT: Fojo, Silvia
; TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying
; TITLE OF INVENTION: Activity and Therapeutic Uses
; FILE REFERENCE: 3806.0505
; CURRENT APPLICATION NUMBER: US/09/846,456
; CURRENT FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/201,280
; PRIOR FILING DATE: 2000-05-02
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 2893
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-846-456-3
```

Query Match:

```
Best Local Similarity 84.9%; Score 1394.4; DB 10; Length 2893;
Matches 1524; Conservative 0; Mismatches 6; Indels 26; Gaps 10;
```

```
QY 16 GCTCCACATGACATCCAGGCGCTGTGG--CTCTTATGGCTCTGCTCTGAGTGT 73
Db 1345 GCCTCCATGACATTCAGGCGCTGTGGCTCTTTCATGCGTCTGCTGAGTGT 1404
QY 74 GATAGAACCATGATGATGAGTACCTGGGCTTGAG--CGTGGCTGGAGATCTGTTGAGTG 132
Db 1405 GATAGAACCATGATGATGAGTACCTGGGCTTGAGCGCTGGCTGGAGATCTGTTGAGTG 1464
QY 133 TAGCATGGAGGGGCTGTGTCAGCTGAATGCTGTATGCAAGTGGTGGAGTCTTGGAA 191
```

1465 TAGCAATGGAGGGGGGCTGTGTCAGCTGATGTCATGATCAGGCTGGTGGAGCTCTTGGAT 1524
192 ATGATGAGCTGGAGGGGGAGAGAGAGTAGGCTGGGGGAGCTCTCTATGACACCTCA 251
1525 ATGATGAGCTGGAGGGGGAGAGAGAGTAGGCTGGGGGAGCTCTCTATGACACCTCA 1584
252 TTCTGGCCAAAACCTCAGCTCAAACTGTGAAGAGCTAAAGTGTGATCTGCGCCCTCAAGGT 311
1585 TTCTGGCCAAAACCTCAGCTCAAACTGTGAAGAGCTAAAGTGTGATCTGCGCCCTCAAGGT 1644
312 GGCTACAAAGGATCTTGTCAAGGTAGAGAGACCTTGTGGCTCCACGCTGACCTCCAGG 371
1645 GGCTACAAAGGATCTTGTCAAGGTAGAGAGACCTTGTGGCTCCACGCTGACCTCCAGG 1704
372 GGCTGCTT-GGCTCTCTCTACGGGTC-GTCCGTGAGTCTCTATGAATCTCCCTCAGGGC 430
1705 GGCTGCTTGGGCTCTCTCTACGGGTCGTCTGAGTCTCTATGAATCTCCCTCAGGGC 1761
431 AGATTCATATTTAGACTCTTCCACAGTTTGACCTGAGTTTGGCCAGAAATAGGTGACATT 490
1762 AGATTCATATTTAGACTCTTCCACAGTTTGACCTGAGTTTGGCCAGAAATAGGTGACATT 1821
491 TAGTTTGTGTGGCTTGAGCAATGACTTAAATATTTAGACATAIGGTGTAGGCTTGCATT 550
1822 TAGTTTGTGTGGCTTGAGTGATGACTTAAATATTTAGAC--ATGGTGTAGGCTTGCATT 1879
551 CCTACTCTTGGCTTTTGTGCCCCCAGGTGTTTGGGTAGTTTGTCTCCCTCAGG 610
1880 CCTACTCTTGGCTTTTGTGCCCCCAGGTGTTTGGGTAGTTTGTCT--CCTCAGG 1938
611 CCAAGGCAACACAGATAAGTTGGAGTCTGGAGTGGGTACATAATTTACACGACTGCAA 670
1939 CCAAGGCAACACAGAGAATGGAGTCTGGAGTGGGTACATAATTTACACGACTGCAA 1998
671 TTCTCTGCTGCATCTTCAAAATGTATACAACTAAATACAAAGTCCCTGTGTATACAC 730
1999 TTCTCTGCTGCATCTTCAAAATGTATACAACTAAATACAAAGTCCCTGTGTATACAC 2058
731 AGGAGGCTGATCAATATATGAATTAAGAGGGGCTGTGCCATATTTCTCTCTTT 790
2059 AGGAGGCTGATCAATATATGAATTAAGAGGGGCTGTGT--CAATATGTCTGTGTTT 2117
791 TGT-----TTGTTTGTCTTTTGTGTTTGTGCTCTCTCTCTCTCAATTA 841
2118 TTCTCTCTTCTTGTGTTTGTCTTTTGTGTTTGTGCTCTCTCTCTCTCAATTA 2177
842 TGAAGAGAGCAAGTAAGATCTTCTCTCGGTCTCTGAGGAGCTGGGAGCTCAGGCT 901
2178 TGAAGAGAGCAAGTAAGATCTTCTCTCGGTCTCTGAGGAGCTGGGAGCTCAGGCT 2237
902 GGAATCTCCAAGGCAAGTGTGCTCTATCAAAATCAAAAGTCCAGCTTGTGGGGGAA 961
2238 GGAATCTCCAAGGCAAGTGTGCTCTATCAAAATCAAAAGTCCAGCTTGTGGGGGAA 2297
962 ACAAAGAGAGCCATTTACCCAGGAGTGTGCGCTTCCCTCACCACCCAGCTTAGGCT 1021
2298 ACAAAGAGAGCCATTTACCCAGGAGTGTGCGCTTCCCTCACCACCCAGCTTAGGCT 2357
1022 TGAAGAGAGCAAGTAAGATCTTGTGCTCTGAGGAGATCTAGCCTTAGA 1081
2358 TGAAGAGAGCAAGTAAGATCTTGTGCTCTGAGGAGATCTAGCCTTAGA 2417
1082 GCTCTCTCTCCCAATPCCCTGCTCGGTGAGGAACTAAAGAGGAGGAGGAGGAGGAGG 1141
2418 GCTCTCTCTCCCAATPCCCTGCTCGGTGAGGAACTAAAGAGGAGGAGGAGGAGGAGG 2477
1142 CGAAGAGAGGATTTAGAGAGCAATTTCCACTGTGTGCTTGGCTGCGGGAGGAGGAGG 1201
2478 CGAAGAGAGGATTTAGAGAGCAATTTCCACTGTGTGCTTGGCTGCGGGAGGAGGAGG 2537
1202 ACTAGAGAGTCTGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1261

2538 ACTAGAGAGTCTGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2597
1262 GCGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1321
2598 GCGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2657
1322 -----CCCCACCCACCCACCTCCGCCAACTCCCTAGATGTGCTGGGGGGTGAAGG 1376
2658 CCCCACCCACCCACCCACCTCCGCCAACTCCGCCAACTCCCTAGATGTGCTGGGGGGTGAAGG 2717
1377 TCGCCCTTTAAGGG 1436
2718 TCGCCCTTTAAGGG 2777
1437 TAAACAGAGGCGGGAAGGG 1496
2778 TAAACAGAGGCGGGAAGGG 2837
1497 CCTCTGGCTCGGTGCAGCCCAATCTATATAAGGAAGTATCCCGGCAAAACCCC 1552
2838 CCTCTGGCTCGGTGCAGCCCAATCTATATAAGGAAGTATCCCGGCAAAACCCC 2893
RESULT 3
US-09-846-456-3/C
; Sequence 3, Application US/09846456
; Patent No. US20020146792A1
; GENERAL INFORMATION:
; APPLICANT: Rosier, Marie
; APPLICANT: Prades, Catherine
; APPLICANT: Lemoine, Cendrine
; APPLICANT: Naudin, Laurent
; APPLICANT: Benefie, Patrice
; APPLICANT: Duverger, Nicolas
; APPLICANT: Brewer, Bryan
; APPLICANT: Remaley, Alan
; APPLICANT: Fojo, Silvia
; TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying
; FILE REFERENCE: 3806.0505
; CURRENT APPLICATION NUMBER: US/09/846,456
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/201,280
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 2893
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-846-456-3
Query Match 6.2%; Score 102.2; DB 10; Length 2893;
Best Local Similarity 65.9%; Pred. No. 1.5e-17;
Matches 164; Conservative 0; Mismatches 83; Indels 2; Gaps 1;
QY 241 ATGCCACTCATCTTGCCCAAACTCAGGTCAAACTGTAAGAGTCTAAATGTAATCTG 300
Db 1820 ATGTCACCTTAT-CTGCCAAAACCTCAGGTCAAACTGTAAGAGTCTAAATGTAATCTG 1761
QY 301 CCCTCAAGTGGCTACAAAAGGTATCTTTGTCAAGGTAGGAGACCTTGTGGCTCCACGT 360
Db 1760 CCCTGAGGATTCATAGAAGACTCAGGACAGACCCGTAGAGAGGCCCAAGAGGCCCTG 1701
QY 361 GCATCTCCAGGGCTTGTGGCTCTTCTACGGGTCTGTCTCTGAGTCTTCTATCAATCTC 420
Db 1700 GAAAGTGCACGTGGAGGC--CACAAAGTCTCTCTACCTTGACAAAGATACCTTTTGTGCCAC 1643
QY 421 CCTTCAGGCGAGATTCATATTAGACTCTTTCAGACTTTGACCTTGTGGCCAGATA 480
Db 1642 CTTGAGGCGAGATTCACATTAGACTCTTTCAGACTTTGACCTTGTGGCCAGATA 1583
QY 481 AGGTGACAT 489

```
Db 1582 AGGTGGCAT 1574
|||||
RESULT 4
US-09-846-456-1/c
; Sequence 1, Application US/09846456
; Patent No. US20020146792A1
; GENERAL INFORMATION:
; APPLICANT: Rosier, Marie
; APPLICANT: Prades, Catherine
; APPLICANT: Lemoine, Cendrine
; APPLICANT: Naudin, Laurent
; APPLICANT: Desefle, Patrice
; APPLICANT: Duverger, Nicolas
; APPLICANT: Brewer, Bryan
; APPLICANT: Remaley, Alan
; APPLICANT: Fojo, Silvia
; TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying It
; FILE REFERENCE: 3806.0505
; CURRENT APPLICATION NUMBER: US/09/846,456
; CURRENT FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/201,280
; PRIOR FILING DATE: 2000-05-02
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 3231
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-846-456-1

Query Match 6.2%; Score 102.2; DB 10; Length 3231;
Best Local Similarity 65.9%; Pred. No. 1.6e-17;
Matches 164; Conservative 0; Mismatches 83; Indels 2; Gaps 1;

Qy 241 ATGCCACCTCATTTGGCCAAACCTCAGTCAAACTGTGAAGAGTCTAAATGTGAATCTG 300
|||||
Db 1820 ATGTCACTTATTCGGCCAAACCTCAGTCAAACTGTGAAGAGTCTAAATGTGAATCTG 1761
|||||
Qy 301 CCTTCAGGTGGTTCACAAAGTATCTTTGTCAAGGTAGGAGACCTTGTGGCCTCCACGT 360
|||||
Db 1760 CCTTGAGGATCTATGAGAGACTCAGGACAGACCGCTAGAGAGCCCAAGCAGCCCTG 1701
|||||
Qy 361 GCATTCAGGCGCTGTGTGGCCCTCTCTTCTAGGGTCTCTCTGAGTCTCTATGAATCTC 420
|||||
Db 1700 GAAGTGCACGTGGAGGC--CACAGGTCTCTCTACCTTACCAAGATACCTTTGAGCCAC 1643
|||||
Qy 421 CCTTCAGGCGCATTCATATTAGACTCTTCACAGTTTGACCTGAGTTTGCCAGATA 480
|||||
Db 1642 CTTCAGGCGCATTCATATTAGACTCTTCACAGTTTGACCTGAGTTTGCCAGATG 1583
|||||
Qy 481 AGGTGACAT 489
|||||
Db 1582 AGGTGGCAT 1574
|||||
RESULT 5
US-09-846-456-4
; Sequence 4, Application US/09846456
; Patent No. US20020146792A1
; GENERAL INFORMATION:
; APPLICANT: Rosier, Marie
; APPLICANT: Prades, Catherine
; APPLICANT: Lemoine, Cendrine
; APPLICANT: Naudin, Laurent
; APPLICANT: Desefle, Patrice
; APPLICANT: Duverger, Nicolas
; APPLICANT: Brewer, Bryan
; APPLICANT: Remaley, Alan
; APPLICANT: Fojo, Silvia
; TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying It
```

```
; TITLE OF INVENTION: Activity and Therapeutic Uses
; FILE REFERENCE: 3806.0505
; CURRENT APPLICATION NUMBER: US/09/846,456
; CURRENT FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/201,280
; PRIOR FILING DATE: 2000-05-02
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 221
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-846-456-4

Query Match 5.5%; Score 91; DB 10; Length 221;
Best Local Similarity 100.0%; Pred. No. 3.7e-15;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1553 GTAATGCGAGCGAGTAGTGGGCGCGGACCGCCGAGCGCGAGCCGACCTTCTCTC 1612
|||||
Db 1 GTAATGCGAGCGAGTAGTGGGCGCGGACCGCCGAGCGCGAGCCGACCTTCTCTC 60
|||||
Qy 1613 CCGGCTCGCGCAGGCGAGGGCGGGAGCTC 1643
|||||
Db 61 CCGGCTCGCGCAGGCGAGGGCGGGAGCTC 91
|||||

RESULT 6
US-09-860-670-255
; Sequence 255, Application US/09850670
; Patent No. US20020165137A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA127P1
; CURRENT APPLICATION NUMBER: US/09/860,670
; CURRENT FILING DATE: 2001-05-21
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 255
; LENGTH: 32190
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-860-670-255

Query Match 2.6%; Score 43.4; DB 9; Length 32190;
Best Local Similarity 66.7%; Pred. No. 0.49;
Matches 62; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 742 TCAATATANTCAATTAAGGGGCGTGGTCCCATATTGTCTGTGTTTGTGTTG 801
|||||
Db 25616 TAAAGAAACACAGAGAGAGAGGCTGATCCCAAGCTACAGGTTTTTTTGTGTTG 25675
|||||
Qy 802 TTCTCTTTTGTGTTTGTGCGCTCCCTCCCT 834
|||||
Db 25676 TTCTCTTTTGTGTTTGTGCGACACCTCGCTCT 25708
|||||

RESULT 7
US-09-860-670-260
; Sequence 260, Application US/09850670
; Patent No. US20020165137A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA127P1
; CURRENT APPLICATION NUMBER: US/09/860,670
; CURRENT FILING DATE: 2001-05-21
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 260
```

```
; LENGTH: 32249
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-860-670-260

Query Match: 2.6%; Score 43.4; DB 9; Length 32249;
Best Local Similarity 66.7%; Pred. No. 0.49;
Matches 62; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 742 TCATATATGAATTAATAAGGGGCTGCTCCCATATGTTCTGTGTTTGTGTTGTTG 801
Db 18834 TAAAGAAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 18893
Qy 802 TTCTTTTGTGTTTGTGCTCCCTCCCTCT 834
Db 18894 TTTGTTGTTGTTTGGAGACAGTCTGCTCT 18926

RESULT 8
US-09-796-692-8332/c
; Sequence 8332, Application US/09796692
; Publication No. US20020198362A
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8332
; LENGTH: 366
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (252)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (255)
; OTHER INFORMATION: n=A,T,C or G
US-09-796-692-8332

Query Match: 2.6%; Score 43; DB 9; Length 366;
Best Local Similarity 65.6%; Pred. No. 0.045;
Matches 61; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
```

```
Qy 742 TCATATATGAATTAATAAGGGGCTGCTCCCATATGTTCTGTGTTTGTGTTGTTG 801
Db 264 TAAAGAAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 205
Qy 802 TTCTTTTGTGTTTGTGCTCCCTCCCTCT 834
Db 204 TTTGTTGTTGTTTGGAGACAGTCTGCTCT 172

RESULT 9
US-09-764-855-60/c
; Sequence 60, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P4110
; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
; LENGTH: 251
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-855-60

Query Match: 2.5%; Score 41.8; DB 10; Length 251;
Best Local Similarity 65.6%; Pred. No. 0.083;
Matches 61; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 772 CCATATGTTCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 831
Db 248 CCCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 189
Qy 832 TCTCAATTTATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 864
Db 188 TTTCAATTTTCAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 156

RESULT 10
US-09-834-975-994
; Sequence 994, Application US/09834975
; Patent No. US20020110815A1
; GENERAL INFORMATION:
; APPLICANT: Brown, Jeffrey
; APPLICANT: Bolt, Andrew
; APPLICANT: Van Huffel, Christophe
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
; TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-016B
; CURRENT APPLICATION NUMBER: US/09/834,975
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,538
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 1045
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 994
; LENGTH: 1974
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-834-975-994

Query Match: 2.5%; Score 41.6; DB 10; Length 1974;
Best Local Similarity 62.5%; Pred. No. 0.31;
Matches 65; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
```

```
Qy 717 TGTGTTTATCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 776
Db 717 TGTGTTTATCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 776
```


Search completed: February 3, 2003, 16:23:44
Job time : 299.737 secs

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.3, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/283,471A
FILING DATE: 04-APR-1999
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/861,233
FILING DATE: 31-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/419,853
FILING DATE: 11-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/483,533
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, James P.
REGISTRATION NUMBER: 28,491
REFERENCE/DOCKET NUMBER: 27373/32742A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 595 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-283-471A-4

Query Match 2.4%; Score 38.8; DB 4; Length 595;
Best Local Similarity 49.2%; Pred. No. 0.21;
Matches 131; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

Qy 1171 CCACCTGTCCTGGCTGGCGGAAACGTGGACTAGAGACTGCGGGGACACCCCGGAGC 1230
Db 99 CCGCAGACACCTGGCGGCTGCTGCGACGCGCGGGGAGGGGCGCGGAGC 158
Qy 1231 CCAGCGTTCGCGGCGCTTAGCGCGCGCGCGCGGCGGGAAGGGGAGCAGACCG 1290
Db 159 CCGCGCGACCCCGCGACCCCGGACCCCGGACCCCGGACCCCGGACCCCGG 218
Qy 1251 CGGACCCCTAAGACACCTGCTGTACCTCCACCCCAACC---CACCCACCTCCCGCCCAAC 1347
Db 219 CGACCCCGGACCCCGGACCCCGGACCCCGGACCCCGGCGGCTCTCGCCCAACG 278
Qy 1348 TCCCTAGATGTCTGCTGGGCGCTGAACGTGCGCGCGGTTAAAGGGCGGCGCGGCTCCA 1407
Db 279 TCCGGGTGCGGACCTGGTGTGCTGGGCTCTGCGCGCGGCTGCGCGCGGCTCGT 338
Qy 1408 CGTCTTCTCTGCTGAGTGAACCT 1433
Db 339 GGGCGCGGAGCGGCGGACCGGCT 364

RESULT 5
US-09-283-471A-4
Sequence 36, Application US/08483533
Patent No. 6172047
GENERAL INFORMATION:
APPLICANT: Roizman, Bernard
APPLICANT: Chou, Joany
TITLE OF INVENTION: Method for Treating Tumorigenic
DISEASES
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago

STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.3, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,533
FILING DATE: 07-MAR-95
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/419,853
FILING DATE: 11-APR-95
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/861,233
FILING DATE: 31-MAR-92
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, James P.
REGISTRATION NUMBER: 28,491
REFERENCE/DOCKET NUMBER: 28097/32742
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1327 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-483-533-36

Query Match 2.4%; Score 38.8; DB 4; Length 1327;
Best Local Similarity 49.2%; Pred. No. 0.33;
Matches 131; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

Qy 1171 CCACCTGTCCTGGCTGGCGGAAACGTGGACTAGAGACTGCGGGGACACCCCGGAGC 1230
Db 540 CCGCAGACACCTGGCGGCTGCTGCGACGCGCGGGGAGGGGCGCGGAGC 599
Qy 1231 CCAGCGTTCGCGGCGCTTAGCGCGCGCGCGCGGCGGGAAGGGGAGCAGACCG 1290
Db 600 CCGCGCGACCCCGGACCCCGGACCCCGGACCCCGGACCCCGGACCCCGG 659
Qy 1291 CGGACCCCTAAGACACCTGCTGTACCTCCACCCCAACC---CACCCACCTCCCGCCCAAC 1347
Db 660 CGACCCCGGACCCCGGACCCCGGACCCCGGACCCCGGACCCCGGACCCCGG 719
Qy 1348 TCCCTAGATGTCTGCTGGGCGCTGAACGTGCGCGGTTAAAGGGCGGCGGCTCCA 1407
Db 720 TCCGGGTGCGGACCTGGTGTGCTGGGCTCTGCGCGCGGCTGCGCGCGGCTCGT 779
Qy 1408 CGTCTTCTCTGCTGAGTGAACCT 1433
Db 780 GGGCGCGGAGCGGCGGACCGGCT 805

RESULT 6
US-09-283-471A-36
Sequence 36, Application US/09283471A
Patent No. 6340673
GENERAL INFORMATION:
APPLICANT: Roizman, Bernard
APPLICANT: Chou, Joany
TITLE OF INVENTION: Method For Treating Tumorigenic Diseases
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago

```

; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60608-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/283,471A
; FILING DATE: 04-APR-1999
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/861,233
; FILING DATE: 31-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/419,853
; FILING DATE: 11-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/483,533
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, James P.
; REGISTRATION NUMBER: 28,491
; REFERENCE/DOCKET NUMBER: 27373/32742A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-5300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1327 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US91-06532-1
; US-09-283-471A-36

```

```

Query Match          2.4%; Score 38.8; DB 4; Length 1327;
Best Local Similarity 49.2%; Pred. No. 0.33;
Matches 131; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

QY 1171 CCACGTGGTCCCTTGGCTGCGCGGACCTGGACTAGAGAGTCTGGCGGCGACGCCCGGAGC 1230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 540 CGCAGAGACACCTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 1290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1231 CCAGCGCTTCCCGCGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 1290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 608 CCGCGCGGACCGCGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1291 CGGACCTTAAGACACCTGCTGTACCTCCACGCCCGGCGGCTTGGCGGCTTGGCGGCTT 1347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 668 CGACCGCGGCGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1348 TCCCTAGATGTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 1407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 728 TCCGGGTGCGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1408 CGTGTCTTCTGCTGAGTGACTGAAC 1433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 788 GGGCGCGGCGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 813
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 7
PCT-US91-06532-1
; Sequence 1, Application PC/TJ59106532
; GENERAL INFORMATION:
; APPLICANT: Roizman, Bernard
; TITLE OF INVENTION: Recombinant Herpes Simplex Viruses
; TITLE OF INVENTION: Vaccines and Methods
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

```

```

; ADDRESSEE: Bicknell
; STREET: Two First National Plaza Suite 2100
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/06532
; FILING DATE: 19910910
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gruber, Lewis S.
; REGISTRATION NUMBER: 30,060
; REFERENCE/DOCKET NUMBER: 27373/8235
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/346-5750
; TELEFAX: 312/984-9740
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1335 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; PCT-US91-06532-1

```

```

Query Match          2.4%; Score 38.8; DB 5; Length 1335;
Best Local Similarity 49.2%; Pred. No. 0.33;
Matches 131; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

QY 1171 CCACGTGGTCCCTTGGCTGCGCGGACCTGGACTAGAGAGTCTGGCGGCGACGCCCGGAGC 1230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 CGCAGAGACACCTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1231 CCAGCGCTTCCCGCGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 1290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 608 CCGCGCGGACCGCGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1291 CGGACCTTAAGACACCTGCTGTACCTCCACGCCCGGCGGCTTGGCGGCTTGGCGGCTT 1347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 668 CGACCGCGGCGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 727
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1348 TCCCTAGATGTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 1407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 728 TCCGGGTGCGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 787
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1408 CGTGTCTTCTGCTGAGTGACTGAAC 1433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 788 GGGCGCGGCGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTTGGCGGCTT 813
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 8
US-09-149-476-208/c
; Sequence 208, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/340,333
; EARLIER FILING DATE: 1997-03-07

```

1	EARLIER APPLICATION NUMBER: 60/043,621	1	EARLIER FILING DATE: 1997-04-11
2	EARLIER FILING DATE: 1997-03-07	2	EARLIER APPLICATION NUMBER: 60/043,674
3	EARLIER APPLICATION NUMBER: 60/040,626	3	EARLIER FILING DATE: 1997-04-11
4	EARLIER FILING DATE: 1997-03-07	4	EARLIER APPLICATION NUMBER: 60/043,315
5	EARLIER APPLICATION NUMBER: 60/040,334	5	EARLIER FILING DATE: 1997-04-11
6	EARLIER FILING DATE: 1997-03-07	6	EARLIER APPLICATION NUMBER: 60/048,974
7	EARLIER APPLICATION NUMBER: 60/040,336	7	EARLIER FILING DATE: 1997-06-06
8	EARLIER FILING DATE: 1997-03-07	8	EARLIER APPLICATION NUMBER: 60/056,886
9	EARLIER APPLICATION NUMBER: 60/040,163	9	EARLIER FILING DATE: 1997-08-22
10	EARLIER FILING DATE: 1997-03-07	10	EARLIER APPLICATION NUMBER: 60/056,877
11	EARLIER APPLICATION NUMBER: 60/047,600	11	EARLIER FILING DATE: 1997-08-22
12	EARLIER FILING DATE: 1997-05-23	12	EARLIER APPLICATION NUMBER: 60/056,889
13	EARLIER APPLICATION NUMBER: 60/047,615	13	EARLIER FILING DATE: 1997-08-22
14	EARLIER FILING DATE: 1997-05-23	14	EARLIER APPLICATION NUMBER: 60/056,893
15	EARLIER APPLICATION NUMBER: 60/047,597	15	EARLIER FILING DATE: 1997-08-22
16	EARLIER FILING DATE: 1997-05-23	16	EARLIER APPLICATION NUMBER: 60/056,630
17	EARLIER APPLICATION NUMBER: 60/047,502	17	EARLIER FILING DATE: 1997-08-22
18	EARLIER FILING DATE: 1997-05-23	18	EARLIER APPLICATION NUMBER: 60/056,878
19	EARLIER APPLICATION NUMBER: 60/047,633	19	EARLIER FILING DATE: 1997-08-22
20	EARLIER FILING DATE: 1997-05-23	20	EARLIER APPLICATION NUMBER: 60/056,666
21	EARLIER APPLICATION NUMBER: 60/047,583	21	EARLIER FILING DATE: 1997-08-22
22	EARLIER FILING DATE: 1997-05-23	22	EARLIER APPLICATION NUMBER: 60/056,872
23	EARLIER APPLICATION NUMBER: 60/047,617	23	EARLIER FILING DATE: 1997-08-22
24	EARLIER FILING DATE: 1997-05-23	24	EARLIER APPLICATION NUMBER: 60/056,862
25	EARLIER APPLICATION NUMBER: 60/047,618	25	EARLIER FILING DATE: 1997-08-22
26	EARLIER FILING DATE: 1997-05-23	26	EARLIER APPLICATION NUMBER: 60/056,637
27	EARLIER APPLICATION NUMBER: 60/047,503	27	EARLIER FILING DATE: 1997-08-22
28	EARLIER FILING DATE: 1997-05-23	28	EARLIER APPLICATION NUMBER: 60/056,903
29	EARLIER APPLICATION NUMBER: 60/047,592	29	EARLIER FILING DATE: 1997-08-22
30	EARLIER FILING DATE: 1997-05-23	30	EARLIER APPLICATION NUMBER: 60/056,888
31	EARLIER APPLICATION NUMBER: 60/047,581	31	EARLIER FILING DATE: 1997-08-22
32	EARLIER FILING DATE: 1997-05-23	32	EARLIER APPLICATION NUMBER: 60/056,879
33	EARLIER APPLICATION NUMBER: 60/047,584	33	EARLIER FILING DATE: 1997-08-22
34	EARLIER FILING DATE: 1997-05-23	34	EARLIER APPLICATION NUMBER: 60/056,880
35	EARLIER APPLICATION NUMBER: 60/047,500	35	EARLIER FILING DATE: 1997-08-22
36	EARLIER FILING DATE: 1997-05-23	36	EARLIER APPLICATION NUMBER: 60/056,894
37	EARLIER APPLICATION NUMBER: 60/047,587	37	EARLIER FILING DATE: 1997-08-22
38	EARLIER FILING DATE: 1997-05-23	38	EARLIER APPLICATION NUMBER: 60/056,911
39	EARLIER APPLICATION NUMBER: 60/047,492	39	EARLIER FILING DATE: 1997-08-22
40	EARLIER FILING DATE: 1997-05-23	40	EARLIER APPLICATION NUMBER: 60/056,636
41	EARLIER APPLICATION NUMBER: 60/047,558	41	EARLIER FILING DATE: 1997-08-22
42	EARLIER FILING DATE: 1997-05-23	42	EARLIER APPLICATION NUMBER: 60/056,874
43	EARLIER APPLICATION NUMBER: 60/047,613	43	EARLIER FILING DATE: 1997-08-22
44	EARLIER FILING DATE: 1997-05-23	44	EARLIER APPLICATION NUMBER: 60/056,910
45	EARLIER APPLICATION NUMBER: 60/047,582	45	EARLIER FILING DATE: 1997-08-22
46	EARLIER FILING DATE: 1997-05-23	46	EARLIER APPLICATION NUMBER: 60/056,864
47	EARLIER APPLICATION NUMBER: 60/047,596	47	EARLIER FILING DATE: 1997-08-22
48	EARLIER FILING DATE: 1997-05-23	48	EARLIER APPLICATION NUMBER: 60/056,631
49	EARLIER APPLICATION NUMBER: 60/047,612	49	EARLIER FILING DATE: 1997-08-22
50	EARLIER FILING DATE: 1997-05-23	50	EARLIER APPLICATION NUMBER: 60/056,845
51	EARLIER APPLICATION NUMBER: 60/047,632	51	EARLIER FILING DATE: 1997-08-22
52	EARLIER FILING DATE: 1997-05-23	52	EARLIER APPLICATION NUMBER: 60/056,892
53	EARLIER APPLICATION NUMBER: 60/047,601	53	EARLIER FILING DATE: 1997-08-22
54	EARLIER FILING DATE: 1997-05-23	54	EARLIER APPLICATION NUMBER: 60/057,761
55	EARLIER APPLICATION NUMBER: 60/043,580	55	EARLIER FILING DATE: 1997-08-22
56	EARLIER FILING DATE: 1997-04-11	56	EARLIER APPLICATION NUMBER: 60/047,595
57	EARLIER APPLICATION NUMBER: 60/043,566	57	EARLIER FILING DATE: 1997-05-23
58	EARLIER FILING DATE: 1997-04-11	58	EARLIER APPLICATION NUMBER: 60/047,599
59	EARLIER APPLICATION NUMBER: 60/043,314	59	EARLIER FILING DATE: 1997-05-23
60	EARLIER FILING DATE: 1997-04-11	60	EARLIER APPLICATION NUMBER: 60/047,588
61	EARLIER APPLICATION NUMBER: 60/043,569	61	EARLIER FILING DATE: 1997-05-23
62	EARLIER FILING DATE: 1997-04-11	62	EARLIER APPLICATION NUMBER: 60/047,585
63	EARLIER APPLICATION NUMBER: 60/043,311	63	EARLIER FILING DATE: 1997-05-23
64	EARLIER FILING DATE: 1997-04-11	64	EARLIER APPLICATION NUMBER: 60/047,586
65	EARLIER APPLICATION NUMBER: 60/043,671	65	EARLIER FILING DATE: 1997-05-23
66	EARLIER FILING DATE		

```

RESULT 9
US-C9-149-476-24/c
: Sequence 24, Application US/09149476
: Patent No. 6420526
: GENERAL INFORMATION:
: APPLICANT: ROSEN ET AL.
: TITLE OF INVENTION: 186 Human Secreted proteins
: FILE REFERENCE: P2003P1
: CURRENT APPLICATION NUMBER: US/09/149,476
: CURRENT FILING DATE: 1998-03-08
: EARLIER APPLICATION NUMBER: PCT/US98/04493
: EARLIER FILING DATE: 1998-03-06
: EARLIER APPLICATION NUMBER: 60/040,162
: EARLIER FILING DATE: 1997-03-07
: EARLIER APPLICATION NUMBER: 60/040,333
: EARLIER FILING DATE: 1997-03-07
: EARLIER APPLICATION NUMBER: 60/038,621
: EARLIER FILING DATE: 1997-03-07
: EARLIER APPLICATION NUMBER: 60/040,621

```



```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041.886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4481 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 163..4099
; US-09-041-886-18

Query Match 2.3%; Score 38.6; DB 4; Length 4481;
Best Local Similarity 52.1%; Pred. No. 0.72;
Matches 86; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 1187 CTGCGGGAACGTGACTAGAGAGTCTGGGGGCGGAGCCCGGAGCCAGCGGTCTCCGGCGC 1246
DB 194 CGCGGTGCGCCACCGAGTCTCCCGCTTCGCGCGAGCCAGGTGGCGGGGTGGCGCTCCG 253
QY 1247 GCTTAGCGCGCGCGCGCGCGGGAAGGGGAGCGCAGACCGCGGAGCCCTAAGACACC 1306
DB 254 TCACGCGCGCGCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 313
QY 1307 TCGTGTACCTCCACCCCGGAGCCCGGAGCCCGGAGCCCGGAGCCCGGAGCCCGGAG 1351
DB 314 COTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 358

RESULT 11
US-08-709-838-1
; Sequence 1, Application: US/08/709838
; Patent No. 5140054
; GENERAL INFORMATION:
; APPLICANT: Loetscher, Marcel
; APPLICANT: Moser, Bernhard
; TITLE OF INVENTION: IP-10/MIG RECEPTOR DESIGNATED CXCR3,
; TITLE OF INVENTION: NUCLEIC ACIDS, AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/709,838
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/709,838
; FILING DATE: 10-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook Esq., David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: TK-96-01A
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1670 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 69..1172
; US-08-829-839-1

;
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1670 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 69..1172
; US-08-709-838-1

Query Match 2.3%; Score 38.2; DB 3; Length 1670;
Best Local Similarity 59.8%; Pred. No. 0.55;
Matches 54; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 714 TCCGTGTTTATCACAGGAGCGCTGATCATATATGAAATTAAGATTAAGAGGGGCGTGGTCC 773
DB 1562 TCTTTTATTTTATGCTCTAAATCTGCTTAAACCTTTCAATAAACAAGATCGTCAGGA 1621
QY 774 CAATATGCTCTGTTTCTTTGTTTCTTTGTTTCTTTGTTTCTTTGTTTCTTTGTTTCT 820
DB 1622 CCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 1668

RESULT 12
US-08-829-839-1
; Sequence 1, Application: US/08829839
; Patent No. 6184358
; GENERAL INFORMATION:
; APPLICANT: Loetscher, Marcel
; APPLICANT: Moser, Bernhard
; APPLICANT: Qin, Shixin
; APPLICANT: Mackay, Charles R.
; TITLE OF INVENTION: IP-10/MIG RECEPTOR DESIGNATED CXCR3,
; TITLE OF INVENTION: ANTIBODIES, NUCLEIC ACIDS, AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/829,839
; FILING DATE: 31-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/709,838
; FILING DATE: 10-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook Esq., David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: TK-96-01A
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1670 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 69..1172
; US-08-829-839-1
```

```
Query Match          2.3%; Score 38.2; DB 4; Length 1670;
Best Local Similarity 59.8%; Pred. No. 0.55;
Matches 64; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 714 TCGTGTATTCACAGCGAGGCTGATCAATAATGAAATTAAGGGGCTGTGCC 773
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1562 TCTTTATTTTATGCTCTAAATCCGCTGCTAAACATTTTCAATAACAAGATCGTCAGGA 1621

QY 774 CATATGTCIGGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 820
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1622 CCATTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTT 1668

RESULT 13
US-09-345-882-1/c
; Sequence 1, Application US/09345882
; Patent No. 6395373
; GENERAL INFORMATION:
; APPLICANT: Bouqueloret, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-7)
; TITLE OF INVENTION: AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
; FILE REFERENCE: GENSET-031A
; CURRENT APPLICATION NUMBER: US/09/345,882
; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-05-30
; PRIOR APPLICATION NUMBER: US 60/111,309
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.pm
; SEQ ID NO 1
; LENGTH: 162450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 72794
; OTHER INFORMATION: 5-124-273 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 88073
; OTHER INFORMATION: 5-127-261 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 90942
; OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 93714
; OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97122
; OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97152
; OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99098
; OTHER INFORMATION: 5-130-257 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99117
; OTHER INFORMATION: 5-130-276 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 103806
; OTHER INFORMATION: 5-131-395 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 106940
; OTHER INFORMATION: 5-128-60 SEQ ID32
```

```
; OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108106
; OTHER INFORMATION: 5-135-355 : polymorphic base insertion of A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108149
; OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTT
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108308
; OTHER INFORMATION: 5-135-357 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108471
; OTHER INFORMATION: 5-136-174 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 134134
; OTHER INFORMATION: 5-140-120 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 134362
; OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 134374
; OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
; FEATURE:
; NAME/KEY: allele
; LOCATION: 145328
; OTHER INFORMATION: 5-143-84 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 146345
; OTHER INFORMATION: 5-143-101 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 150329
; OTHER INFORMATION: 5-145-24 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 160031
; OTHER INFORMATION: 5-148-352 : polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 72771..72817
; OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
; FEATURE:
; NAME/KEY: allele
; LOCATION: 72771..72817
; OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
; FEATURE:
; NAME/KEY: allele
; LOCATION: 88050..88096
; OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
; FEATURE:
; NAME/KEY: allele
; LOCATION: 88050..88096
; OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
; FEATURE:
; NAME/KEY: allele
; LOCATION: 90819..90865
; OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
; FEATURE:
; NAME/KEY: allele
; LOCATION: 90819..90865
; OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
; FEATURE:
; NAME/KEY: allele
; LOCATION: 93690..93736
; OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
```



```
FEATURE:
NAME/KEY: allele
LOCATION: 93690..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID72
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106966
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108084..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:
```

Query Match 2.3%; Score 37.8; DB 4; Length 162450;
Best Local Similarity 52.2%; Pred. No. 8.4;
Matches 84; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
QY 586 TCACAAATGATACAAACTAATACAGCTCCTGTCGTTTTTATCAGGAGGCTGATCAA 745
DB 109593 TAAACATTTTACTATATTAATAGAGCCCATTTTATTTCCCTCCCTGTTGCT 109534
QY 746 TATAATGAATTAAGGGGCTGTCATATGTCGTTGTTGTTGTTGTTGTTGTTGTT 805
DB 109593 CTGATACTGTGCCAAGAGTACAGTAAATGATTTTATTTTATTTTATTTT 109474
QY 806 TTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 846
DB 109473 TTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 109433

RESULT 14
US-08-018-977C-4/c
; Sequence 4, Application US/08018577C
; Patent No. 5686601
; GENERAL INFORMATION:
; APPLICANT: Weber, Peter C.
; TITLE OF INVENTION: DNA Sequences Encoding Mutant Antiviral
; TITLE OF INVENTION: Regulatory Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Pennsylvania State University
; STREET: 113 Technology Center
; CITY: University Park
; STATE: PA
; COUNTRY: US
; ZIP: 16802-7000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS 6.22
; SOFTWARE: Microsoft Word for Windows 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/018,977C
; FILING DATE: 18-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/726,071
; FILING DATE: 05-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas J. Monahan
; REGISTRATION NUMBER: 29,835
; REFERENCE/DOCKET NUMBER: 91-1039/2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (814) 865-6277
; TELEFAX: (814) 865-3591
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 936 bases
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-018-977C-4
```

Query Match 2.3%; Score 37.6; DB 1; Length 936;
Best Local Similarity 54.3%; Pred. No. 0.59;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
QY 1212 CTGGCGGCGAGCCGCGAGCCGAGCGCTCCCGCGCTTTAGCGCGGCGGCGGCGG 1271
DB 785 CTGGAGGACCGCGCGCTCTCTGTCGACAGAGCGGCGGCGGCGGCGGCGGCGG 726
QY 1272 GCGAAGGAGCGAGACCGCGGAGCCCTTAAGACACCTGTGTACCTCCACCCACCCA 1331
DB 725 GAGGGGCGGCGCGCGGAGGCGCGCTGCCACCTCCACCGCGGCGGCGGCGGCGG 666

QY 1332 CCCACCTCCCGCACTCC 1351
DB 665 CGCGCGACCGTCGACGCGC 646
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 12:46:48 : Search time 418.41 seconds
(without alignments)
8245.850 Million cell updates/sec

Title: US-09-596-141C-3_COPY_1_1532

Perfect score: 1532
Sequence: 1 gaattccctgtgggtc.....accggaatctataaaaggaa 1532

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1532	100.0	1643	22	AAF24681
2	1532	100.0	1643	22	AAF24703
3	1387.6	90.6	18399	22	AAF92831
4	1374.4	89.7	2910	24	AAD37267
5	1374.4	89.7	3231	24	AAD37265
6	1035.4	67.6	1197	24	ABL58400
7	102.2	6.7	2910	24	AAD37267
8	102.2	6.7	3231	24	AAD37265
9	97	6.3	1643	22	AAF24681

C 10	97	6.3	1643	22	AAF24703	Nucleotide sequenc
C 11	92.6	6.0	18399	22	AAF92831	Human ABC1 genomic
C 12	51.4	3.4	474	24	ABN73478	Bovine embryonic g
C 13	45.6	3.0	6116	24	ABN80129	Human chemically m
C 14	44.2	2.9	456	22	AAI80464	Human polynucleoti
C 15	43.8	2.9	385	22	AAI83199	Human polynucleoti
C 16	43.8	2.9	400	22	AAI83631	Human polynucleoti
C 17	43.4	2.8	401	22	AAI85424	Human polynucleoti
C 18	43.4	2.8	772	22	AAH03293	Human cDNA clone (
C 19	43.4	2.8	2339	22	AAH16964	Human cDNA sequenc
C 20	43.4	2.8	10119	22	AAK87559	Human immune/haema
C 21	43.4	2.8	14983	22	AAK73075	Human immune/haema
C 22	43.4	2.8	14983	22	AAK87542	Human immune/haema
C 23	43.4	2.8	32190	22	AAI62927	Human genomic DNA
C 24	43.4	2.8	32249	22	AAI62932	Human immune/haema
C 25	43.4	2.8	39068	22	AAK71820	Human immune/haema
C 26	43.4	2.8	39068	22	AAK73078	Human immune/haema
C 27	43.4	2.8	39068	22	AAK85294	Human immune/haema
C 28	43.4	2.8	39068	22	AAK87544	Human immune/haema
C 29	43.4	2.8	39110	22	AAK71525	Human immune/haema
C 30	43.4	2.8	39110	22	AAK73087	Human immune/haema
C 31	43.4	2.8	39110	22	AAK87555	Human immune/haema
C 32	43.4	2.8	45300	22	AAK73079	Human immune/haema
C 33	43.4	2.8	45300	22	AAK87547	Human immune/haema
C 34	43.2	2.8	421	22	AAI89084	Human polynucleoti
C 35	43.2	2.8	466	22	AAI82085	Human polynucleoti
C 36	43.2	2.8	595	23	ABV16167	Human prostate exp
C 37	43	2.8	470	22	AAH08333	Human cDNA clone (
C 38	42.6	2.8	620	21	AAQ59594	Human secreted pro
C 39	42.4	2.8	386	22	AAI92146	Human polynucleoti
C 40	42.4	2.8	389	22	AAI85397	Human polynucleoti
C 41	42.4	2.8	5690	22	AAK81236	Human immune/haema
C 42	42.4	2.8	5690	22	AAK81239	Human immune/haema
C 43	42.2	2.8	429	22	AAI87584	Human polynucleoti
C 44	42.2	2.8	24259	22	AAK66691	Tumour suppressor
C 45	42	2.7	387	22	AA-63088	Human polynucleoti

ALIGNMENTS

RESULT 1

AAF24681
ID AAF24681 standard; DNA; 1543 BP.

XX AAF24681;

XX 20-APR-2001 (first entry)

DE Nucleotide sequence of the 5' flanking region of the human ABC1 gene.

KW Human; adenosine triphosphate binding cassette protein 1; ABC1;

KW apolipoprotein-mediated mobilisation; cholesterol; tangier disease;

KW chromosome 9p22-9q31; heart disease; hypercholesterolemia;

KW atherosclerosis; cholesterol transport; ss.

OS Homo sapiens.

PN WO200078972-A2.

PD 28-DEC-2000.

XX 16-JUN-2000; 2000WC-US16765.

XX 18-JUN-1999; 99US-0140264.

PR 14-SEP-1999; 99US-0153872.

PR 19-NOV-1999; 99US-0165573.

PA (CVTH-) CV THERAPEUTICS INC.

XX Lawn Rv, Wade D, Garvin M;

XX WPI; 2001-13782/14.

DR

XX		Adenosine triphosphate (ATP) binding cassette (ABC) polynucleotide,
PI		useful for the development of agents for the treatment of heart disease
PI		and other disorders associated with hypercholesterolemia and
PI		atherosclerosis -
PS		Claim 1; Page 143-144; 215pp; English.
XX		The present sequence represents the 5' flanking region of the human
CC		adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABC1
CC		resides in cell membranes and utilizes ATP hydrolysis to transport a wide
CC		variety of substrates across the p-lasma membrane. ABC1 is a pivotal
CC		protein in the apolipoprotein-mediated mobilisation of intracellular
CC		cholesterol stores. ABC1 is defective in Tangier disease, a genetic
CC		disorder characterised by abnormal HDL-cholesterol metabolism. The ABC1
CC		gene is localised to chromosome 9q22-9q31. The ABC1 genes and proteins
CC		are useful for developing pharmaceutical agents for the treatment of
CC		heart disease and other disorders associated with hypercholesterolemia
CC		and atherosclerosis. The genes are useful for developing screening assays
CC		to screen for compounds that regulate the expression of genes associated
CC		with cholesterol transport. The genes and proteins are also useful for
CC		are also useful as diagnostic indicators of cardiovascular disease and
CC		other disorders associated with hypercholesterolemia.
XX		Sequence 1643 BP: 370 A; 413 C; 457 G; 403 T; 0 other;
SQ		
	Query Match:	100.0%; Score 1532; DB 22; Length 1643;
	Rest Local Similarity	100.0%; Pred. No. 0;
	Matches 1532; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1	GAATTCCTGCTGGTGCACATGCACCTCCAGGGCGCTGTGGCTCTCTCATTGGGTC 60
DB	1	GAATTCCTGCTGGTGGCTCCATGCACCTCCAGGGCGCTGTGGCTCTCTCATTGGGTC 60
QY	61	TGTCTCGAGTGTTGATAGAACCACTGATGTGAGTAGCTGGCGCTGAGCGTGGCGCTGGAGA 120
DB	61	TGTCTCGAGTGTTGATAGAACCACTGATGTGAGTAGCTGGCGCTGAGCGTGGCGCTGGAGA 120
QY	121	TCCTGTACTGTAGCATGAGGGGCTTTGTCCAGCTGAATGTCCTATGCCAGGTGGTGGG 180
DB	121	TCCTGTACTGTAGCATGAGGGGCTTTGTCCAGCTGAATGTCCTATGCCAGGTGGTGGG 180
QY	181	ASTCTCGAATATGATGGAGCTGGAGTGGGAAGACAAGTACGCTTGGCGAGGCTCTC 240
DB	181	AGTCTCGAATATGATGGAGCTGGAGTGGGAAGACAAGTACGCTTGGCGAGGCTCTC 240
QY	241	ATGCCAOCCTATTTCTGGCCAAAACACTCAGGTCAAACCTGTGAAGAGCTAAATGTGAATCTG 300
DB	241	ATGCCAOCCTATTTCTGGCCAAAACACTCAGGTCAAACCTGTGAAGAGCTAAATGTGAATCTG 300
QY	301	CCCTTAAGGTGGCTACAAAGGTATCTTTGTCAAGTAGGAGACCCTGTGGCGCTCCAGGT 360
DB	301	CCCTTAAGGTGGCTACAAAGGTATCTTTGTCAAGTAGGAGACCCTGTGGCGCTCCAGGT 360
QY	361	GCATCTCCAGGGCGCTGCTGGCGCTCTTCTACGGGCTGTGCTCCGAGTC-TCTATGAATCTC 420
DB	361	GCATCTCCAGGGCGCTGCTGGCGCTCTTCTACGGGCTGTGCTCCGAGTC-TCTATGAATCTC 420
QY	421	CCTTCAGGGCAGATTCATA-TTAGACTCTTACAGTCTTTCACAGTTTACCCTGAGTTTTGCCCAAGATA 480
DB	421	CCTTCAGGGCAGATTCATA-TTAGACTCTTACAGTCTTTCACAGTCTTTCACAGTATGGTCTGA 540
QY	481	AGGTGACATTTAGT-TTGTGGCTGTGATGAAGTAACTAAATATTTAGACATATGGTGTGTA 540
DB	481	AGGTGACATTTAGT-TTGTGGCTGTGATGAAGTAACTAAATATTTAGACATATGGTGTGTA 540
QY	541	GSCCTGCAATCTCTACTCTTGCCCTTTTCTTTTGGCCCCCTCCAGTCTTTGGGCTAGTTTGGCT 600
DB	541	GSCCTGCAATCTCTACTCTTGCCCTTTTCTTTTGGCCCCCTCCAGTCTTTTGGGCTAGTTTGGCT 600
QY	601	CCCCCTACAGCCAAGGCAACAGATAAGTTGGAGGTCTCTGGAGTGGCTACATAATTTTAC 660
DB	601	CCCCCTACAGCCAAGGCAACAGATAAGTTGGAGGTCTCTGGAGTGGCTACATAATTTTAC 660

KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
 KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
 KW atherosclerosis; cholesterol transport; ss.

XX Homo sapiens.

XX WO200078971-A2.

XX 28-DEC-2000.

XX 16-JUN-2000; 2000WO-US15591.

XX 18-JUN-1999; 99US-0140264.

XX 14-SEP-1999; 99US-0153872.

XX 19-NOV-1999; 99US-0166573.

XX (CVTH-) CV THERAPEUTICS INC.

XX (UNIW) UNIV WASHINGTON.

XX Lawn RM, Wade D, Oram JF, Garvin M;

XX WPI; 2001-137811/14.

XX Adenosine triphosphate (ATP) binding cassette protein (ABC) 1
 PT polynucleotides and polypeptides, useful for treatment of heart disease
 PT and other disorders associated with hypercholesterolemia and
 PT atherosclerosis.

XX Disclosure; Page 138-139; 21pp; English.

XX The present sequence represents the 5' flanking region of the human
 CC adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABC1
 CC resides in cell membranes and utilizes ATP hydrolysis to transport a wide
 CC variety of substrates across the plasma membrane. ABC1 is a pivotal
 CC protein in the apolipoprotein-mediated mobilisation of intracellular
 CC cholesterol stores. ABC1 is defective in Tangier disease, a genetic
 CC disorder characterised by abnormal HDL-cholesterol metabolism. The ABC1
 CC gene is localised to chromosome 9q22-9q31. The ABC1 genes and proteins
 CC are useful for developing pharmaceutical agents for the treatment of
 CC heart disease and other disorders associated with hypercholesterolemia
 CC and atherosclerosis. The genes are useful for developing screening assays
 CC to screen for compounds that regulate the expression of genes associated
 CC with cholesterol transport. The genes and proteins are also useful for
 CC as also useful as diagnostic indicators of cardiovascular disease and
 CC other disorders associated with hypercholesterolemia.

XX Sequence 1643 bp; 370 A; 413 C; 457 G; 403 T; 0 other;

Query Match 100.0%; Score 1532; DB 22; Length 1643;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCCTGCTGGTCCACATGCACTCCAGGGCTGCTTGGCTCTCTATGGTC 60

DB 1 GAATTCCTGCTGGTCCACATGCACTCCAGGGCTGCTTGGCTCTCTATGGTC 60

QY 61 TGTCTCAGTGTGATGAACCACTGATGTCAGTACCTTGGCTTCCAGGCTGGCCCTGAGA 120

DB 61 TGTCTCAGTGTGATGAACCACTGATGTCAGTACCTTGGCTTCCAGGCTGGCCCTGAGA 120

QY 121 TCTGTGAGCTGATGAGGGGGCTTGCAGTGAATGTCGTATGAGGTGGTGGG 180

DB 121 TCTGTGAGCTGATGAGGGGGCTTGCAGTGAATGTCGTATGAGGTGGTGGG 180

QY 181 AGTCTGCAATATGATGAGTGGAGTGGGAAGAAGATAGGCTTGGGCGAGCTCTCTC 240

DB 181 AGTCTGCAATATGATGAGTGGAGTGGGAAGAAGATAGGCTTGGGCGAGCTCTCTC 240

QY 241 ATGCCACCTCATCTGCGCAAACTCAGGTCAAACTGTSAAGAGTCTAAATGTGAATCTG 300

DB 241 ATGCCACCTCATCTGCGCGCAAACTCAGGTCAAACTGTSAAGAGTCTAAATGTGAATCTG 300

QY 301 CCTTCAGAGTGGCTACAAAGTATCTTTGTCAAGTATGAGACACCTTGTGGCCTCCACGT 360

DB 301 CCTTCAGAGTGGCTACAAAGTATCTTTGTCAAGTATGAGACACCTTGTGGCCTCCACGT 360

QY 361 GCACCTCCAGGGCTGCTGGGCTCTCTACAGGCTCTGCTGAGTCTTCTATGAATCTC 420

DB 361 GCACCTCCAGGGCTGCTGGGCTCTCTACAGGCTCTGCTGAGTCTTCTATGAATCTC 420

QY 421 CCTTCAGGGCAGATTCATATTTAGACTCTTCAAGCTTTCACCTGAGTTTGGCCAGAA 480

DB 421 CCTTCAGGGCAGATTCATATTTAGACTCTTCAAGCTTTCACCTGAGTTTGGCCAGAA 480

QY 481 AGGTGACATTTAGTTGGCTTGGCTGATGAATGACTTAAATATTTAGACATATGGTGTGA 540

DB 481 AGGTGACATTTAGTTGGCTTGGCTGATGAATGACTTAAATATTTAGACATATGGTGTGA 540

QY 541 GGCTTGCATCTCTAGTCTTGGCTTTTGGCCCTCCAGTCTTTTGGGTAGTTTGGT 600

DB 541 GGCTTGCATCTCTAGTCTTGGCTTTTGGCCCTCCAGTCTTTTGGGTAGTTTGGT 600

QY 601 CCCCTACAGCCAAAGGCAACAGATAAGTTGGAGTCTGAGTGGCTACATAATTTTAC 660

DB 601 CCCCTACAGCCAAAGGCAACAGATAAGTTGGAGTCTGAGTGGCTACATAATTTTAC 660

QY 661 AGGACTGCAATTCCTGGCTGGCTGCACTTCAAAATGTATACAACTAAATCAAGTCCGTG 720

DB 661 AGGACTGCAATTCCTGGCTGGCTGCACTTCAAAATGTATACAACTAAATCAAGTCCGTG 720

QY 721 TTTTATPCACAGGAGGCTGATCAATATATGAATTAAGTAAAGGGGGTGGTCCCATATTC 780

DB 721 TTTTATPCACAGGAGGCTGATCAATATATGAATTAAGTAAAGGGGGTGGTCCCATATTC 780

QY 781 TTTTGTCTTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 840

DB 781 TTTTGTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 840

QY 841 ATGAAGAGAGCAGTATAGATGTTCTCTCGGCTCTCTGAGGACCTTGGGAGCTCAGGC 900

DB 841 ATGAAGAGAGCAGTATAGATGTTCTCTCGGCTCTCTGAGGACCTTGGGAGCTCAGGC 900

QY 901 TGGGAATCTCCAAAGGCTAGTCTGCTATCAAAATCAAAAGTCCAGTCTTGTGGGGGA 960

DB 901 TGGGAATCTCCAAAGGCTAGTCTGCTATCAAAATCAAAAGTCCAGTCTTGTGGGGGA 960

QY 961 AAACAAAGAGCCCATATCCAGAGGACTGTCGGCTTCCCTCACCAGCTAGGCC 1020

DB 961 AAACAAAGAGCCCATATCCAGAGGACTGTCGGCTTCCCTCACCAGCTAGGCC 1020

QY 1021 TTTCAGAGGAACAAAGACAGACAAATGATTTGGCTCTCTGAGGAGATTCAGCCTAG 1080

DB 1021 TTTCAGAGGAACAAAGACAGACAAATGATTTGGCTCTCTGAGGAGATTCAGCCTAG 1080

QY 1081 AGCTCTCTCTCCCCCAATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1140

DB 1081 AGCTCTCTCTCCCCCAATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1140

QY 1141 CGGAAAGCAGATTTAGAGAGCAAAATCCACTGTTGCCCTTGGCTGCCGGGAAGGTG 1200

DB 1141 CGGAAAGCAGATTTAGAGAGCAAAATCCACTGTTGCCCTTGGCTGCCGGGAAGGTG 1200

QY 1201 GACTAGAGAGTCTGGGGCGCAGCCCGAGCCGCTTCCCGCGGCTTTAGCCCGGCG 1260

DB 1201 GACTAGAGAGTCTGGGGCGCAGCCCGAGCCGCTTCCCGCGGCTTTAGCCCGGCG 1260

QY 1261 GGCCCGGCGGGGAAGGGGACGACGCCGCGGACCCCTTAAGACACCTTGTCTACCTTCCA 1320

DB 1261 GGCCCGGCGGGGAAGGGGACGACGCCGCGGACCCCTTAAGACACCTTGTCTACCTTCCA 1320

QY 1321 CCCCAACCCCAACCTTCCCGGCACTCCCTAGTGTGTTGGGGCGGCTGAAGCTGCG 1380

DB 1321 CCCCAACCCCAACCTTCCCGGCACTCCCTAGTGTGTTGGGGCGGCTGAAGCTGCG 1380

QY 1381 CGGTTTAAAGGCGGGGCGGCTCCAGTGTCTTCTGCTGAGTGAAGTGAATATAA 1440


```
FI protein_bind 2051..2059
FT /*tag= ak
FT /bound_moiety= "GATA"
FI 2104..2111
FT /*tag= al
FT /bound_moiety= "SOX5"
FI 2114..2152
FT /*tag= am
FT /bound_moiety= "SYR/HFH/HNF3beta"
FI 2221..2228
FT /*tag= an
FT /bound_moiety= "KZF1"
FI 2234..2249
FT /*tag= ao
FT /bound_moiety= "IK2/NFkappaB/CREL"
FI 2259..2272
FT /*tag= ap
FT /bound_moiety= "LMO2COM/GATA"
FI 2289..2306
FT /*tag= aq
FT /bound_moiety= "KZF1/SRY"
FI 2313..2318
FT /*tag= ar
FT /bound_moiety= "PPAR"
FI 2321..2326
FT /*tag= as
FT /bound_moiety= "PPAR"
FI 2335..2342
FT /*tag= at
FT /bound_moiety= "KZF1"
FI 2361..2384
FT /*tag= au
FT /bound_moiety= "HNF3beta/SRY/EV11"
FI 2426..2433
FT /*tag= av
FT /bound_moiety= "KZF1"
FI 2442..2451
FT /*tag= aw
FT /bound_moiety= "AP4"
FI 2455..2466
FT /*tag= ax
FT /bound_moiety= "SRY"
FI 2491..2498
FT /*tag= ay
FT /bound_moiety= "STAT"
FI 2524..2534
FT /*tag= az
FT /bound_moiety= "STAT/PPAR"
FI 2536..2541
FT /*tag= ba
FT /bound_moiety= "PPAR"
FI 2589..2600
FT /*tag= bb
FT /bound_moiety= "AP2"
FI 2610..2617
FT /*tag= bc
FT /bound_moiety= "KZF1"
FI 2634..2648
FT /*tag= bd
FT /bound_moiety= "LMO2COM/MYOD/E47"
FI 2657..2672
FT /*tag= be
FT /bound_moiety= "RREB1"
FI 2680..2698
FT /*tag= bf
FT /bound_moiety= "KZF1/CMYB"
FI 2728..2740
FT /*tag= bg
FT /bound_moiety= "SPI/GC"
FI 2743..2757
FT /*tag= bh
FT /bound_moiety= "USE/NMYC/ARNT"
FI 2758..2773
```

```
FT /*tag= bi
FT /bound_moiety= "NFE2A1"
FI 2774..2787
FT /*tag= bj
FI 2794..2806
FT /*tag= bk
FT /bound_moiety= "GC/SP1/KZFI"

Query Match 89.7%; Score 1374.4; DB 24; Length 2910;
Best Local Similarity 97.9%; Prec. No. 0;
Matches 1504; Conservative 0; Mismatches 6; Indels 26; Gaps 10:

Qy 16 GGTCCACATGCACATCCAGGCGCTTGG--CTCTTCTATGGCTCTGTCTGAGTGT 73
Db |||||
Qy 1345 GCTCCACATGCACATCCAGGCGCTTGGGCGCTTCTATGGCTC:GTCCIGAGTGT 1404
Db |||||

Qy 74 GATAGAACCTGATGTGAGTACCTGGGCTTGCAG--CGTGGGCTGGAGATCCTGTGTGATG 132
Db |||||
Qy 1405 GATAGAACCTGATGTGAGTACCTGGGCTTGGAGCGGTGGCTGGAGATCCTGTGTGACIG 1464
Db |||||

Qy 133 TAGCATGGAGGGGCTTGT--CAGCTGAATGTCTGTATGTCAGGTGGTGGAGTCTTGGAAAT 191
Db |||||
Qy 1465 TAGCATGGAGGGGCTTGTGACCTGATGCTGTCATGCAGGTGGTGGAGTCTTGTGGAAT 1524
Db |||||

Qy 192 ATGATGGAGCTGGAGTGGGAAGAGAAGTAGCTTTGGGCGAGCTCTCTATGCCACCTCA 251
Db |||||
Qy 1525 ATGATGGAGCTGGAGTGGGAAGAGAAGTAGCTTTGGGCGAGCTCTCTATGCCACCTCA 1584
Db |||||

Qy 252 TTCTGGCCAAAACCTCAGGTCAACCTGTGAAGAGTCTAAATGTGAATCTGCCCTTCAAGGT 311
Db |||||
Qy 1585 TTCTGGCCAAAACCTCAGGTCAACCTGTGAAGAGTCTAAATGTGAATCTGCCCTTCAAGGT 1644
Db |||||

Qy 312 GCCTACAAAGGTATCTTGTCAAGGTAGGAGACCTTTGGGCTCCACGTGCACTTCCAGG 371
Db |||||
Qy 1645 GCCTACAAAGGTATCTTGTCAAGGTAGGAGACCTTTGGGCTCCACGTGCACTTCCAGG 1704
Db |||||

Qy 372 GCTGCTT--GGCTCTCTCPACGGGCTCTGCTGAGTCTTCTATGAATCTCCCTTCAGGCG 430
Db |||||
Qy 1705 GCTGCTTGGGCTCTCTCPACGGGCTCTGCTGAGTCTTCTATGAAT--CGTTCAGGCG 1761
Db |||||

Qy 431 AGATTCATATTTAGACTCTTCAACAGTTTGACCTGAGTTTGGCCAGATAAAGTGCACAT 490
Db |||||
Qy 1762 AGATTCATATTTAGACTCTTCAACAGTTTGACCTGAGTTTGGCCAGATAAAGTGCACAT 1821
Db |||||

Qy 491 TAGTTTGTGGCTTGATGAATGACTTAAATATTTAGACATATGGTGTAGCCCTGCATT 550
Db |||||
Qy 1822 TAGTTTGTGGCTTGATGAATGACTTAAATATTTAGAC--ATGGTGTGTAGCCCTGCATT 1879
Db |||||

Qy 551 CCTACTCTTGGCTTTTGTGGGCTTCCAGTGTGTTTGGGTAGTTTGTGCTCCCTACAG 610
Db |||||
Qy 1880 CCTACTCTTGGCTTTTGTGGGCTTCCAGTGTGTTTGGGTAGTTTGTGCTCCCTACAG 1938
Db |||||

Qy 611 CCAAGGCAACAGATAAGTTTGGAGTCTGGAGTGGCTACATAAATTTTACAGACTGCAA 670
Db |||||
Qy 1939 CCAAGGCAACAGATAAGTTTGGAGTCTGGAGTGGCTACATAAATTTTACAGACTGCAA 1998
Db |||||

Qy 671 TTCTCTGGCTGCACCTTCAACAATGTATACAACTAAATAACAAGTCTGTGTTTATATCAG 730
Db |||||
Qy 1999 TTCTCTGGCTGCACCTTCAACAATGTATACAACTAAATAACAAGTCTGTGTTTATATCAG 2058
Db |||||

Qy 731 AGGAGGCTGTCAATATATAATGAATTAAGAGGGGCTGTCCCATATTTGTCTGTGTT 790
Db |||||
Qy 2059 AGGAGGCTGTCAATATATAATGAATTAAGAGGGGCTGTCCCATATTTGTCTGTGTT 2117
Db |||||

Qy 791 TTG-----TTTGTGTTCTCTTTTGTGTTTGTGTTTGGCTCCTCCTCATTTA 841
Db |||||
Qy 2118 TTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 2177
Db |||||

Qy 842 TGAAGAGAGCAGTAAGATGTTCTCTCGGGTCTCTGAGGGACCTGGGAGCTCAGGCT 901
Db |||||
Qy 2178 TGAAGAGAGCAGTAAGATGTTCTCTCGGGTCTCTGAGGGACCTGGGAGCTCAGGCT 2237
Db |||||
```


QY 502 GGAATCTCCAGCAGTAGTGGCTATCAAAATCAAGTCCAGGTTTGGGGGAA 961
 Db 2238 GGGAACTCCAAAGCAGTAGTGGCTATCAAAATCAAGTCCAGGTTTGGGGGAA 2297
 QY 962 AACAAAGCAGCCATACCCAGAGGACITCTCCGCTTCCCTCACCCAGCCTAGGCT 1021
 Db 2258 AACAAAGCAGCCATACCCAGAGGACITCTCCGCTTCCCTCACCCAGCCTAGGCT 2357
 QY 1022 TTGAAGGAAACAAAGACAAAGAAATGATTTGGCTTCTGAGGAGATTCAGCCTAGA 1081
 Db 2358 TTGAAGGAAACAAAGACAAAGAAATGATTTGGCTTCTGAGGAGATTCAGCCTAGA 2417
 QY 1082 GCTCTCTCCCCCAATCCCTCCCTCCGCTGAGAACTAACAAAGGAAAGAAAAATG 1141
 Db 2418 GCTCTCTCCCCCAATCCCTCCCTCCGCTGAGAACTAACAAAGGAAAGAAAAATG 2477
 QY 1142 CGGAAGCAGGATTTAGAGGAGGAAATTCACCTGGTCCCTTGGCTCCGCGGAACGTGG 1201
 Db 2478 CGGAAGCAGGATTTAGAGGAGGAAATTCACCTGGTCCCTTGGCTCCGCGGAACGTGG 2537
 QY 1202 ACTAGAGAGTCTGGCGCGCAGCCCGAGCCAGCGCTTCCCGCGCTTTAGCGCGCGG 1261
 Db 2538 ACTAGAGAGTCTGGCGCGCAGCCCGAGCCAGCGCTTCCCGCGCTTTAGCGCGCGG 2597
 QY 1262 GCCCGCGGGGAGGAGGAGCAGACCGCGGACCTTAAGACACCTGTGTACCCCTCCAC 1321
 Db 2598 GCCCGCGGGGAGGAGGAGCAGACCGCGGACCTTAAGACACCTGTGTACCCCTCCAC 2657
 QY 1322 -----CCCCACCCACCCCTCCCGCACTCCCTAGATGTGCTGGCGGCGCTGAAGC 1376
 Db 2658 CCCACCCACCCACCCCTCCCGCACTCCCTAGATGTGCTGGCGGCGCTGAAGC 2717
 QY 1377 TGGCCCGCTTAAGGGCGGGGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1436
 Db 2718 TGGCCCGCTTAAGGGCGGGGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2777
 QY 1437 TAAACAGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1496
 Db 2778 TAAACAGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2837
 QY 1497 COTCTGCGCTCGGTGAGCGGCAATCTATAAAGGAA 1532
 Db 2838 COTCTGCGCTCGGTGAGCGGCAATCTATAAAGGAA 2873

RESULT 5
 ID AAD37265

AC AAD37265; standard; DNA; 3231 Bp.

XX AAD37265;

XX 21-AUG-2002 (first entry)

DE Human ABC1 transcription regulatory DNA #1.

DE Human: ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
 KW cholesterol metabolism; hypercholesterolemia; antisense therapy; ds.

OS Homo sapiens.

XX WO200183746-A2.

PN 08-NOV-2001.

XX 02-MAY-2001; 2001WO-EP05486.

XX 02-MAY-2000; 2000US-201280P.

XX (AVET) AVENTIS PHARMA SA.

XX Rosier-Montus M, Prades C, Lemoine C, Naudin L, Benefle P;

PI Brewer B, Duverger N, Remaley A, Santamarina-Fojo S;

XX

DR WPI; 2002-154404/20.
 XX Isolated nucleic acid useful for modifying the ATP-binding cassette 1
 PT (ABC1) and screening for candidate modulatory compounds or substances
 PT
 XX
 PS Claim 1; Page 130-131; 152pp; English.
 XX
 CC The invention relates a nucleic acid which is capable of regulating the
 CC transcription of human ATE-binding cassette 1 (ABC1) gene, which is a
 CC casual gene for pathologies linked to a dysfunctioning of cholesterol
 CC metabolism, including diseases such as atherosclerosis. Polynucleotides
 CC that are capable of modulating the transcription of the ABC1 gene. They
 CC are used in antisense therapy. Compositions comprising sequences of the
 CC invention are used to treat hypercholesterolemia and atherosclerosis.
 CC The present sequence is human ABC1 transcription regulating DNA.
 XX
 SQ Sequence 3231 Bp; 809 A; 773 C; 876 G; 773 T; 0 other;

Query Match 89.7%; Score 1374.4; DB 24; Length 3231;
 Best Local Similarity 97.9%; Pred. No. 0;
 Matches 1504; Conservative 0; Mismatches 6; Indels 26; Gaps 10;

QY 16 GGCTCCACATGACATTCACAGGCGCTCTGG--CTCTTCTATGGTCTGTCCAGTGT 73
 Db 1345 GCTCCACAIGACATTCACAGGCGCTCTGGCGCTCTTCTATGGCTGTCTGTGAGTGT 1404
 QY 74 GATACAACACATGATGATGATACCTGGGCTTGAG--CGTGGCGCTGAGATCCTGTGACTG 132
 Db 1405 GATACAACACATGATGATGATACCTGGGCTTGAGCGCTGGGCTGGAGATCCTGTGACTG 1464
 QY 133 TAGCATGGAGGGGCGCTTGT--CAGCTCAATGTCTGTATGAGGTGGTGGAGTCTCTGGAAT 191
 Db 1465 TAGCATGGAGGGGCGCTTGTGCAGCTGAATGCTGCATGCAGGTGGTGGAGTCTCTGGAAT 1524
 QY 192 ATGATGGAGCTGGAGTGGGAGAGAGTAGCTTGGGGCAGCTCTCTCATGCCACCTCA 251
 Db 1525 ATGATGGAGCTGGAGTGGGAGAGAGTAGCTTGGGGCAGCTCTCTCATGCCACCTCA 1584
 QY 252 TCTGGCCAAAACCTCAGGTCAAACGTGAAGAGTCTAAATGTGAATCTGCCCTTCAAGT 311
 Db 1585 TCTGGCCAAAACCTCAGGTCAACACTGTGAAGAGTCTAAATGTGAATCTGCCCTTCAAGT 1644
 QY 312 GGCTACAAAGGTATCTTTGTCAAGGTAGAGACCTTTGGGCTCCACGTTGCACTTCCAGG 371
 Db 1645 GGCTACAAAGGTATCTTTGTCAAGGTAGAGACCTTTGGGCTCCACGTTGCACTTCCAGG 1704
 QY 372 GGCTGCT--GGCCTCTCTACGGGTCTGTCTGAGTCTTCTATGAATCTCCCTTCAGGCG 430
 Db 1705 GGCTGCTGGGCTCTCTACGGGTCTGTCTGAGTCTTCTATGAATCTCCCTTCAGGCG 1761
 QY 431 ACATTCATATTTAGACTCTTTCACAGTCTTACAGTCTTGGCTGGCTTGGCCAGAAATAGCTGACATT 490
 Db 1762 ACATTCATATTTAGACTCTTTCACAGTCTTACAGTCTTGGCTGGCTTGGCCAGAAATAGCTGACATT 1821
 QY 491 TASTTTGTGGCTTCATGAATGACTTAATAATTTAGACATATGGTGTAGGCGCTTGCATT 550
 Db 1822 TATGTTGTGGCTTCATGAATGACTTAATAATTTAGAC--ATGGTGTGTAGGCGCTTGCATT 1879
 QY 551 CCTACTCTTGGCTTTTCTTTCCTCCCTCCAGTCTTTGGGTAGTCTTGGCTCCCTACAG 610
 Db 1880 CCTACTCTTGGCTTTTCTTTCCTCCCTCCAGTCTTTGGGTAGTCTTGGCTCCCTACAG 1938
 QY 611 CCAAGGCAACACAGATAGTGTGAGGTCTGAGGTGGCTGACATTAATTTACAGCACTGCAA 670
 Db 1939 CCAAGGCAACACAGAGAGTGTGAGGTGGCTGACATTAATTTACAGCACTGCAA 1996
 QY 671 TTCTCTGGCTGCACCTTCACAAATGTATACAAACTAAATACAAAGTCTGTGTTTATCAC 730
 Db 1999 TTCTCTGGCTGCACCTTCACAAATGTATACAAACTAAATACAAAGTCTGTGTTTATCAC 2058
 QY 731 AGGGAGGCTGATCAATATTAATGAATTAAGAGGGGCTGGTCCCATATTTGTTCTGTGTT 790

XX (AVEI) AVENTIS PHARMA SA.
 PA Rosier-Montus M, Prades C, Lemoine C, Naudin L, Benefle P;
 PI Brewer B, Duverger N, Remaley A, Santamarina-Fojo S;
 XX WPI; 2002-154404/20.
 DR
 XX
 XX Isolated nucleic acid useful for modifying the ATP-binding cassette 1
 PT (ABC1) and screening for candidate modulatory compounds or substances
 PT
 XX
 XX Claim 1; Page 130-131; 152pp; English.
 XX
 XX The invention relates a nucleic acid which is capable of regulating the
 CC transcription of human ATP-binding cassette 1 (ABC1) gene, which is a
 CC casual gene for pathologies linked to a dysfunctioning of cholesterol
 CC metabolism, including diseases such as atherosclerosis. Polynucleotides
 CC of the invention are used to screen candidate molecules or substances
 CC that are capable of modulating the transcription of the ABC1 gene. They
 CC are used in antisense therapy. Compositions comprising sequences of the
 CC invention are used to treat hypercholesterolemia and atherosclerosis.
 CC The present sequence is human ABC1 transcription regulating DNA.
 XX
 XX Sequence 3231 BP; 809 A; 773 C; 876 G; 773 T; 0 other;
 SQ
 Query Match 6.7%; Score 102.2; DB 24; Length 3231;
 Best Local Similarity 65.9%; Pred. No. 1.6e-17;
 Matches 164; Conservative 0; Mismatches 83; Indels 2; Gaps 1;
 QY 241 ATGCCACCTCATTCTGGCCAAACTCAGTCAAACTGTGAAGAGTCTAAATGTAATCTG 300
 DB 1820 ATGTCACCTTATTCTGGCCAAACTCAGTCAAACTGTGAAGAGTCTAAATGTAATCTG 1761
 QY 301 CCCTTCAGGTGGTACAAAGGTATCTTTGTCAAGGTAGGAGACCTTGTGGCCCTCCAGC 360
 DB 1760 CCCTGAGGAGTTCATAGAAGACACAGGACAGCCCTGAGAGGCCCAAGCAGCCCTG 1701
 QY 361 GCACCTTCAGGCGCTGTGGCCCTCTTCTAGGGTCTCTCTGAGTCTTCATGATGATCTC 420
 DB 1700 GAAGTGCACGTGGAGGC--CACAGGTCTCTACCTTGACAAAGATACCTTTGTAGCCAC 1643
 QY 421 CCTTCAGGCGCAGATTCATATTAGACTCTTCACACTTTGACCTGAGTTTGGCCAGAATA 480
 DB 1642 CTTGAAGGCGCAGATTCACATTTAGACTCTTCACAGTTTGACCTGAGTTTGGCCAGAATG 1583
 QY 481 AGGTGACAT 489
 DB 1582 AGGTGGCAT 1574
 RESULT 9
 AAF24681/c
 ID AAF24681 standard; DNA; 1643 BP.
 XX
 XX AAF24681;
 AC
 XX
 XX 20-APR-2001 (first entry)
 DT
 XX
 XX Nucleotide sequence of the 5' flanking region of the human ABC1 gene.
 XX
 XX Human; adenosine triphosphate binding cassette protein 1; ABC1;
 KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
 KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
 KW atherosclerosis; cholesterol transport; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200078972-A2.
 PN
 XX
 XX 28-DEC-2000.
 PD
 XX
 XX 16-JUN-2000; 2000WO-US16765.

XX 18-JUN-1999; 99US-0140264.
 PR 14-SEP-1999; 99US-0153872.
 PR 19-NOV-1999; 99US-0166573.
 XX
 XX (CVTH-) CV THERAPEUTICS INC.
 XX
 XX Lawn RM, Wade D, Garvin M;
 PI WPI; 2001-137812/14.
 DR
 XX
 XX Adenosine triphosphate (ATP) binding cassette (ABC) polynucleotide,
 PT useful for the development of agents for the treatment of heart disease
 PT and other disorders associated with hypercholesterolemia and
 PT atherosclerosis -
 XX
 XX Claim 1; Page 143-144; 215pp; English.
 XX
 XX The present sequence represents the 5' flanking region of the human
 CC adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABC1
 CC resides in cell membranes and utilises ATP hydrolysis to transport a wide
 CC variety of substrates across the plasma membrane. ABC1 is a pivotal
 CC protein in the apolipoprotein-mediated mobilisation of intracellular
 CC cholesterol stores. ABC1 is defective in Tangier disease, a genetic
 CC disorder characterised by abnormal HDL-cholesterol metabolism. The ABC1
 CC gene is localised to chromosome 9q22-9q31. The ABC1 genes and proteins
 CC are useful for developing pharmaceutical agents for the treatment of
 CC heart disease and other disorders associated with hypercholesterolemia
 CC and atherosclerosis. The genes are useful for developing screening assays
 CC to screen for compounds that regulate the expression of genes associated
 CC with cholesterol transport. The genes and proteins are also useful for
 CC are also useful as diagnostic indicators of cardiovascular disease and
 CC other disorders associated with hypercholesterolemia.
 XX
 XX Sequence 1643 BP; 370 A; 413 C; 457 G; 403 T; 0 other;
 SQ
 Query Match 6.3%; Score 97; DB 22; Length 1643;
 Best Local Similarity 62.8%; Pred. No. 3.2e-16;
 Matches 154; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
 QY 241 ATGCCACCTCATTCTGGCCAAACTCAGTCAAACTGTGAAGAGTCTAAATGTAATCTG 300
 DB 489 ATGTCACCTTATTCTGGCCAAACTCAGTCAAACTGTGAAGAGTCTAAATGTAATCTG 430
 QY 301 CCCTTCAGGTGGTACAAAGGTATCTTTGTCAAGGTAGGAGACCTTGTGGCCCTCCAGC 360
 DB 429 CCCTGAGGAGATTCATAGAAGACTCAGGACAGCCCGTAGAAGGCCAAGCAGGCC 370
 QY 361 GCACCTTCAGGCGCTGTGGCCCTCTTCTAGGGTCTCTCTGAGTCTTCATGATGATCTC 420
 DB 369 TGAAGTGCACGTGGAGGCCACAGGTCTCTCTACCTTGACAAAGATACCTTTGTAGCCAC 310
 QY 421 CCTTCAGGCGCAGATTCATATTAGACTCTTCACACTTTGACCTGAGTTTGGCCAGAATA 480
 DB 309 CTTGAAGGCGCAGATTCACATTTAGACTCTTCACAGTTTGACCTGAGTTTGGCCAGAATG 250
 QY 481 AGGTGACAT 489
 DB 249 AGGTGGCAT 241
 RESULT 10
 AAF24703/c
 ID AAF24703 standard; DNA; 1643 BP.
 XX
 XX AAF24703;
 AC
 XX
 XX 20-APR-2001 (first entry)
 DT
 XX
 XX Nucleotide sequence of the 5' flanking region of the human ABC1 gene.
 XX
 XX Human; adenosine triphosphate binding cassette protein 1; ABC1;
 KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
 KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
 KW atherosclerosis; cholesterol transport; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200078972-A2.
 PN
 XX
 XX 28-DEC-2000.
 PD
 XX
 XX 16-JUN-2000; 2000WO-US16765.

KW ch-omosome 9q22-9q31; heart disease; hypercholesterolemia;
KW atherosclerosis; cholesterol transport; ss.

OS Homo sapiens.

PN WO200078971-A2.

PD 28-DEC-2000.

PF 16-JUN-2000; 2000WO-US16591.

PR 18-JUN-1999; 99US-0140264.

PR 14-SEP-1999; 99US-0153872.

PR 19-NOV-1999; 99US-0166573.

XX (CVTH-) CV THERAPEUTICS INC.

PA (UNIW) UNIV WASHINGTON.

PI Lawn RM, Wade D, Oram JF, Garvin M;

DR WPI; 2001-137811/14.

XX Adenosine triphosphate (ATP) binding cassette protein (ABC) i

PT polynucleotides and polypeptides, useful for treatment of heart disease

PT and other disorders associated with hypercholesterolemia and

XX atherosclerosis -

PS Disclosure; Page 138-139; 211pp; English.

XX The present sequence represents the 5' flanking region of the human
CC adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABC1
CC resides in cell membranes and utilizes ATP hydrolysis to transport a wide
CC variety of substrates across the plasma membrane. ABC1 is a pivotal
CC protein in the apolipoprotein-mediated mobilization of intracellular
CC cholesterol stores. ABC1 is defective in Tangier disease, a genetic
CC disorder characterised by abnormal HDL-cholesterol metabolism. The ABC1
CC gene is localised to chromosome 9q22-9q31. The ABC1 genes and proteins
CC are useful for developing pharmaceutical agents for the treatment of
CC heart disease and other disorders associated with hypercholesterolemia
CC and atherosclerosis. The genes are useful for developing screening assays
CC to screen for compounds that regulate the expression of genes associated
CC with cholesterol transport. The genes and proteins are also useful for
CC are also useful as diagnostic indicators of cardiovascular disease and
CC other disorders associated with hypercholesterolemia.

XX Sequence 1643 BP; 370 A; 413 C; 457 G; 403 T; 0 other;

Query Match 6.3%; Score 97; DB 22; Length 1643;
Best Local Similarity 51.8%; Pred. No. 3.2e-16;
Matches 154; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 241 ATGCCACCTCATCTGCGCAAACTCAGTCAAACTGGAAGTCTAAATGTGAATCTG 300

DB 489 ATGTCACCTTATCTGCGCAAACTCAGTCAAACTGGAAGTCTAAATGTGAATCTG 430

QY 301 CCTTCAAGTGTCTACAAGTATCTTTGTCGAAGTGTAGAGACCTTGTGGCTCCACGT 360

DB 429 CCTTCAAGTGTCTACAAGTATCTTTGTCGAAGTGTAGAGACCTTGTGGCTCCACGT 370

QY 361 GCACCTCCAGGGCTCTCTGGCTCTCTACCGGTCTCTGCTCTCTCTCTCTCTCTCT 420

DB 369 TGAAGTGTACGTGGAGGCGACAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 310

QY 421 CCTTCAAGTGTCTACAAGTATCTTTGTCGAAGTGTAGAGACCTTGTGGCTCCACGT 480

DB 309 CTGGAAGGCGAGATTCACATTAGACTCTCTACAGTGTACAGTGTACAGTGTACAGT 250

QY 481 AGGTGACAT 489

DB 249 AGGTGACAT 241

RESULT 11

AAF92831/C
ID AAF92831 standard; DNA; 183999 BP.

XX AC AAF92831;

XX DT 17-MAY-2001 (first entry)

XX DE Human ABC1 genomic DNA.

XX KW High density lipoprotein-cholesterol; HDL-C; cardiovascular; ABC1; ds.

OS Homo sapiens.

PN WO200115676-A2.

XX PD 08-MAR-2001.

XX PF 01-SEP-2000; 2000WO-IB01492.

XX PR 01-SEP-1999; 99US-0151977.

XX PR 15-MAR-2000; 2000US-0526193.

XX PR 23-JUN-2000; 2000US-0213958.

XX PA (UYBR-) UNIV BRITISH COLUMBIA.

XX PA (XENO-) XENON GENETICS INC.

XX PI Hayden MR, Brooks-Wilson AR, Pinstone SN, Clee SM;

XX WPI; 2001-244356/25.

XX Treating a lower than normal high density lipoprotein-cholesterol
CC (HDL-C) level, a higher than normal triglyceride level, or a
CC cardiovascular disease, by administering a compound that modulates LXR-
CC or RXR-mediated transcriptional activity -

XX Claim 8; Fig 1; 317pp; English.

XX The present invention relates to a method for treating a patient
CC diagnosed as having a lower than normal high density
CC lipoprotein-cholesterol (HDL-C) level, a higher than normal
CC triglyceride level, or a cardiovascular disease, involving
CC administering a compound that modulates LXR- or RXR-mediated
CC transcriptional activity or ABC1 expression or activity.
CC The LXR gene product may be used in an assay to identify
CC compounds useful for the treatment of a disease or condition selected a
CC lower than normal HDL cholesterol level, a higher than normal
CC triglyceride level, and a cardiovascular disease.

XX Sequence 183999 BP; 49549 A; 37944 C; 41170 G; 54950 T; 386 other;

Query Match 6.0%; Score 92.6; DB 22; Length 183999;
Best Local Similarity 63.1%; Pred. No. 5.4e-14;
Matches 159; Conservative 0; Mismatches 90; Indels 3; Gaps 1;

QY 241 ATGCCACCTCATCTGCGCAAACTCAGTCAAACTGGAAGTCTAAATGTGAATCTG 300

DB 27714 ATGTCACCTTATCTGCGCAAACTCAGTCAAACTGGAAGTCTAAATGTGAATCTG 27655

QY 301 CCTTCAAGTGTCTACAAGTATCTTTGTCGAAGTGTAGAGACCTTGTGGCTCCACGT 360

DB 27654 CCTTCAAGTGTCTACAAGTATCTTTGTCGAAGTGTAGAGACCTTGTGGCTCCACGT 27595

QY 361 GCACCTCCAGGGCTCTCTGGCTCTCTACCGGTCTCTGCTCTCTCTCTCTCTCTCT 417

DB 27594 CCTTCAAGTGTCTACAAGTATCTTTGTCGAAGTGTAGAGACCTTGTGGCTCCACGT 27535

QY 418 CTGCTTCAAGGCGAGATTCACATTAGACTCTCTACAGTGTACAGTGTACAGTGTACAGT 477

DB 27534 CACCTTCAAGGCGAGATTCACATTAGACTCTCTACAGTGTACAGTGTACAGTGTACAGT 27475

QY 478 ATAGGTGACAT 489

DB 27474 ATAGGTGACAT 27463

Qy 816 TTGTGGCCTCCTTCTCTCAATTTATGAAAGAGAGAGTAA 858
||| | | | | | | | | | | | | | | | |
Db 49 TTTTGGTTCAGAGCGGTCAAGTTAAGTTGAAATCTCCTAAG 7

Search completed: February 3, 2003, 16:28:38
Job time : 933.553 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 16:39:45 ; Search time 2898.81 Seconds
(without alignments)
8559.210 Million cell updates/sec

Title: us-09-596-141c-3_copy_1_1532

Perfected score: 1532
Sequence: 1 gaatccttggtggtgc.....agccgaatctataaaaggaa 1532

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

ESI:*

```

1:  em_estba:*
2:  em_esthum:*
3:  em_estin:*
4:  em_estmd:*
5:  em_estov:*
6:  em_estpl:*
7:  em_estro:*
8:  em_btc:*
9:  gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_othr:*
26: em_gss_pro:*
27: em_gss_rod:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
C 1	245.4	16.0	289 10	AW748338 RC6-BT025
2	211.4	7.3	467 9	AA527406 ng37c02.s
3	60.8	4.0	1201 17	AL05981 Drosophil
4	54.8	3.6	298 17	AL295898 Tetraodon
5	53.8	3.5	844 17	AL05652 Drosophil
6	53	3.5	237 9	AU222489 AU222489

C 7	52.4	3.4	910	17	CNS0060N	AL065629 Drosophil
8	51.2	3.3	832	17	CNS015XC	AL106036 Drosophil
C 9	51	3.3	681	17	CNS02FE9	AL194922 Tetraodon
C 10	50.6	3.3	925	17	CNS0091P	AL053013 Drosophil
C 11	50.4	3.3	844	17	CNS0052P	AL056652 Drosophil
C 12	50.4	3.3	870	17	CNS005E2	AL094271 Drosophil
C 13	48.4	3.2	1004	14	BQ944204	BQ944204 AGENCOURT
C 14	48	3.1	900	17	AG081217	AG081217 Pan trogl
C 15	47.6	3.1	1009	17	CNS010FW	AL058882 Drosophil
C 16	47.4	3.1	1059	14	BQ680479	BQ680479 AGENCOURT
C 17	47.4	3.1	1101	17	CNS017Y2	AL108536 Drosophil
C 18	47.2	3.1	477	9	AL513813	AL513813 AL513913
C 19	47.2	3.1	925	17	CNS0091P	AL053013 Drosophil
C 20	47.2	3.1	942	17	AG042878	AG042878 Pan trogl
C 21	47	3.1	658	12	BF968907	BF968907 602270874
C 22	47	3.1	843	17	CNS005S1	AL059666 Drosophil
C 23	47	3.1	973	17	CNS0080F	AL052232 Drosophil
C 24	47	3.1	997	17	CNS006DN	AL065132 Drosophil
C 25	46.8	3.1	431	9	AL513947	AL513947 AL513947
C 26	46.8	3.1	976	17	CNS0068V	AL063041 Drosophil
C 27	46.6	3.0	978	17	AG030617	AG030617 Pan trogl
C 28	46.6	3.0	1101	17	CNS0083U	AL056399 Drosophil
C 29	46.2	3.0	527	9	AL514325	AL514325 AL514325
C 30	46.2	3.0	875	17	AG043475	AG043475 Pan trogl
C 31	46.2	3.0	1136	14	BQ943816	BQ943816 AGENCOURT
C 32	46.2	3.0	1417	14	BM610236	BM610236 AGENCOURT
C 33	46	3.0	758	17	AG073529	AG073529 Pan trogl
C 34	45.8	3.0	203	10	AW150511	AW150511 xg32h07.x
C 35	45.8	3.0	471	10	BE503803	BE503803 hv83a05.x
C 36	45.8	3.0	523	9	AL514015	AL514015 AL514015
C 37	45.8	3.0	766	17	AG041031	AG041031 Pan trogl
C 38	45.8	3.0	901	17	CNS0760F	AL430789 clone XBA
C 39	45.8	3.0	932	17	CNS00720	AL066742 Drosophil
C 40	45.8	3.0	987	17	CNS017LB	AL108185 Drosophil
C 41	45.8	3.0	997	14	BQ953024	BQ953024 AGENCOURT
C 42	45.6	3.0	223	13	BG995916	BG995916 MR4-HT119
C 43	45.6	3.0	235	10	BE012563	BE012563 122176 MA
C 44	45.6	3.0	456	9	AL554951	AL554951 AL554951
C 45	45.6	3.0	526	9	AL657934	AL657934 AL657934

ALIGNMENT'S

RESULT 1
AW748338/C AW748338 289 bp mRNA linear EST 28-APR-2000
LOCUS RC6-BT0252-271099-012-cl0 BT0252 Homo sapiens CDNA, mRNA sequence.
DEFINITION
ACCESSION AW748338
VERSION AW748338.1 GI:7663270
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
1 (bases 1 to 289)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homitidae; Homo.
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.P.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL
MEDLINE
COMMENT
Proc. Natl. Acad. Sci. U.S.A. 97 (7). 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?c1=RC6&L2=RC6-BT0252-271099-012-cl0&t3=1999-10-27&t4=1)

Seq primer: puc 18 forward
 High quality sequence start: 146
 High quality sequence stop: 289.

FEATURES

Location/Qualifiers
 ..289
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT0252"
 /dev_stage="Adult"

/note="Organ: breast; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (J.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 94 a 70 c 51 g 74 t

ORIGIN

Query Match 15.0%; Score 245.4; DB 10; Length 289;
 Best Local Similarity 96.9%; Pred. No. 1.6e-38;
 Matches 282; Conservative 0; Mismatches 6; Indels 3; Gaps 3;
 QY 678 GTGCACCTTCACAAAGTATACAACTAAATACAAAGTCCTGTTTATATACAGGGAGG 737
 DB 289 GTGCACCTTCACAAAGTATACAACTAAATACAAAGTCCTGTTTATATACAGGGAGG 230
 QY 738 CTGATCAATATATGAATATAAAGGGGCTGTCCTCATATTTCTGTGTTTGTGTTG 797
 DB 229 CTGATCAATATATGAATATAAAGGGGCTGCT-CCATATTTCTGTGTTTGTGTTG 171
 QY 798 TTGTTTCTTTTGTGTTTGTGCTCTCTCTCTCAATTTATGAAGAGAGCAGTAA 857
 DB 170 TTGTTTCTTTTGTGTTTGT-CCCTCTCTCTCTCAATTTATGAAGAGAGCAGTAA 112
 QY 858 GATGTCCTCTCGGGTCCTCTCAGGAGACCTGGGAGCTCAGGCTGGGAATCTCCAAGGCA 917
 DB 111 GATGTCCTCTCGGGTCCTCTCAGGAGACCTGGGAGCTCAGGCTGGG-ATCTCCAAGGCA 53
 QY 918 GTAGGTCGCTATCAAAATCAAGTCCAGTTTG*GGGGGAAACAAAA 968
 DB 52 GTAGGTCGCTATCAAAATCAAGTCCAGTTTGTGGGGGAAACAAAA 2

RESULT 2

AA527406 467 bp mRNA linear EST 21-AUG-1997
 LOCUS ng37c02.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:935962 3',
 DEFINITION mRNA sequence.

ACCESSION AA527406

VERSION AA527406.1 GI:2269475

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 467)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.

, Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arraying: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.lnl.gov/bbrp/image/image.html
 Insert Length: 1057 Std Error: 0.00
 Seq primer: -40ml3 fwd. EI from Amersham
 High quality sequence stop: 380.

FEATURES

Location/Qualifiers
 ..467
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="NCI_CGAP_Co3"
 /sex="pooled"
 /tissue_type="colon"

/lab_host="DH10B"
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization."

BASE COUNT 72 a 118 c 166 g 111 t

ORIGIN

Query Match 7.3%; Score 111.4; DB 9; Length 467;
 Best Local Similarity 99.1%; Pred. No. 4.4e-12;
 Matches 112; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 816 TTGTGGCTCTCTCTCAATTTATGAAGAGAGCAGTAAGATGTTCTCTCGGGTCC 875
 DB 1 TTGTGGCTCTCTCTCAATTTATGAAGAGAGCAGTAAGATGTTCTCTCGGGTCC 60
 QY 876 TCTGAGGACCTGGGGAGCTCAGGCTGGGAATCTCCAGGAGTAGTCGCT 926
 DB 61 TCTGAGGACCTGGGGAGCTCAGGCTGGGAATCTCCAGGAGTAGTCGCT 113

RESULT 3

CNS015W3 1201 bp DNA linear GSS 26-JUL-1999
 LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
 DEFINITION BACN15123 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL105981

VERSION AL105981.1 GI:5619397

KEYWORDS GSS.

SOURCE Drosophila melanogaster.

ORGANISM Drosophila melanogaster.

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1201)

REFERENCE Genoscope.

AUTHORS Direct Submission

TITLE Drosophila

JOURNAL

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk - . This Drosophila melanogaster BAC library (Dros-BAC) was made by Alain Billaut at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES

Location/Qualifiers
 ..1201
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_lib="BACN15123"
 /clone_lib="DrosBAC"


```

Qy 1117 AACTATACARAGGAAARAAATTCGCGAAGCAGGATTIAGAGGAAGCAAAATTCACCTG 1176
Db 301 AAAAAAAAAAAGAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 360

Qy 1177 GTGCCCTTGCGTGGCGGGAACGTGGACGTAGAGATCTCGCGCGAGCCCGAGCCGACGCG 1236
Db 361 SASVSSACSVYAGSSMAAGASCBGCAANWCACGCCSSSSASRSRMMKSSGSSCASSSCS 420

Qy 1237 CTCTCCCGCGCTTTAGCCCGCGCGCCCGCGCGGGAAGGAGGACGACCGCGGACC 1296
Db 421 VGSSSCSAMSSCCCVVSSCGMASSCCCGGMSASSSSSSGGSVSVSGRAVGRGCMCC 480

Qy 1297 CTAAGACACCTGCTACCTCCACCCACCCACCCACCCACCCACCCACCCACCCACCC 1356
Db 481 S-----WCCCMCSMCMCCSVSSVCAVCSGVSRSVAVGCVSGVRVGGSSRGRAGRS 535

Qy 1357 GTGTGTGGCGCGGTGAACGTGCGCGCTTAAGGGCGGCGCGCGCGCGCGCGCTTTC 1416
Db 536 SGRGGSSVSGVSSSVGMCACSSASSVSCSBSSVASSVSSBVGRCRGVGGV 595

Qy 1417 TGCTGACTGACTGACTACATAACAGAGAGCGCGGAGCGCGCGGAGGAGGAGAGCA 1476
Db 596 GGGRVSSCGSSGSSSSGCGSVRSRSGSANGVRVGGSGRGGGGRGGRGANARAD 655

Qy 1477 CA 1478
Db 656 DA 557

RESULT 6
AU222489 237 bp mRNA linear EST 03-APR-2002
LOCUS AU222489
DEFINITION AU222489 Rice callus (2001); Oryza sativa (japonica cultivar-group)
cDNA clone C10354, mRNA sequence.
ACCESSION AU222489
VERSION AU222489.1 GI:15008101
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group).
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 237)
Sasaki, T. and Yamamoto, K.
Rice cDNA from callus (2001)
Unpublished (2001)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasakieabr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
PROJECT = 'RGP', Location/Qualifiers
..237
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="C10354"
/clone_lib="Rice callus (2001)"
/tissue_type="callus"
BASE COUNT 69 a 24 c 38 g 1c2 t 4 others
ORIGIN
Query Match 3.5%; Score 53; DB 9; Length 237;
Best Local Similarity 59.7%; Prec. No. 2;
Matches 86; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 716 CTGTCTTTTAAACACAGGAGGCGTGAATATAATGAATATAAGGAGGCGGTGTCTCCA 775
Db 17 CTTATTTTAAATCGAGNGTGTNTNGCAAAAAAAGGCGCGCCCTT 76

```

```

Qy 776 TATGTCTCGTGTTTTTTTGTGTGTTTCTTTTTTTTGTGGCTCTTCTCTC 835
Db 77 TTTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAACTCTGA 136

Qy 836 AATTATCAAGAGAAGCAGTAAGA 859
Db 137 CACTGGGAAGGGAGAGTTTGA 160

RESULT 7
CNS0060N/c 910 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence 17 end of BAC #
DEFINITION BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION A.065629
VERSION A.065629.1 GI:4944698
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 910)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
3P 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org/The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Marmoser in Pletier de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
Location/Qualifiers
..910
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR14J21"
/clone_lib="RPCI-98"
/note="end : T7"
BASE COUNT 202 a 63 c 112 g 198 t 335 others
ORIGIN
Query Match 3.4%; Score 52.4; DB 17; Length 910;
Best Local Similarity 20.5%; Prec. No. 1.6;
Matches 48; Conservative 104; Mismatches 82; Indels 0; Gaps 0;

Qy 1118 AACTAACAAAGAAAGAAAAAATTCGGGAAGCAGGATTTAGAGGAAGCAATTCATGG 1177
Db 752 AASMARVAAVVARVASAAVAVASAAVAVASVAVSVASASRVMASAGSSSSSSSASS 693

Qy 1178 TGCCCTTGCTCGCGGGAACGTGGACTAGAGAGTCTCGCGCGACGCCCGCCAGCCGCGC 1237
Db 652 SASSSASSASVSVSCRSAVSVASAAASAAASASASASRAVAVAVASVAVASVSSSV 633

Qy 1238 TTCGCCCGCGTCTTAGCGCGCGCGCGCGCGGAGGAGCGAGACCGCGGACCC 1297
Db 632 SCSSSRCCSSASSASASSGSSSSSSSSSSCGCGSCASCASSSSRSRSCGCGSVGVS 573

Qy 1298 TAAGACACTCTGTACCTCCACCCACCCACCCACCCACCCACCCACCCACCC 1351

```



```

Db 373 CTBSSGTSBSTS 361

RESULT 12
CNS006EZ/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACRI3E15 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL064271.1 GI:4944345
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster.
ORGANISM
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 870)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
Bp 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osada and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
Location/Qualifiers
1..870
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACRI3E15"
/clone_lib="RPCI-98"
/note="end : TET3"

BASE COUNT 256 a 90 c 87 g 127 t 310 others
ORIGIN
Query Match 3.3%; Score 50.4; DB 17; Length 870;
Best Local Similarity 28.0%; Pred. No. 4.1;
Matches 67; Conservative 74; Mismatches 98; Indels 0; Gaps 0;

QY 1174 CTGTGTCCTTGCTGCGGGAAGTGGAGTCTAGAGAGTCTGCGGCGAGCCCGAGCCCA 1233
D 1174 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 809 CSSGTGTTCTSSSTTTCCGCKSSGGTSSCBFTTGGSCSTCCSSSTCTCTCTTCGG 750
QY 1234 GCGTTTCCGCGGTCTTAGCCCGCGGCGCGGCGGGAAGAGGAGCAGACCGCGG 1233
D 1234 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 749 GSCSTSSSTTCKKSSSGSKANKYKBSBGCSVYVGSASMSGCGKASAGKRKS 690
QY 1294 ACCTAGACACCTGCTGTAGCTTCCAGCCCGACCCACCCACCTCCCGCACTCCCTA 1353
D 1294 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 689 ASSGCGSSGSGSGSGYACCCMCCSCCYCCCGCCYCYCCSCCTCTTTCGGBSC 630
QY 1354 GATGTGCTGTGGCGGTGACGTCGCCCTTTAAAGGCGGCGCGCCCGCCACCTGCG 1412
D 1354 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 GCTSBKSGKSSSGNGTGTCTCCCTCTCCSCYCGCGCCGCTCCSCYSCS 571

RESULT 13
BQ944204/c
LOCUS
DEFINITION
Pan troglodytes male lymphoblast DNA, clone PTB-077C21.F.
ACCESSION
AG081217.1 GI:15633019
KEYWORDS
GSS.
SOURCE
Pan troglodytes male lymphoblast DNA, clone PTB-077C21.F.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Primates; Catarrhini; Hominidae; Pan.

BASE COUNT 12c a 319 c 374 g 152 t 38 others
ORIGIN
Query Match 3.2%; Score 48.4; DB 14; Length 1004;
Best Local Similarity 52.0%; Pred. No. 9.6;
Matches 91; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 1177 GTGCCCTTGCTGCGGGAAGTGGAGTCTAGAGAGTCTGCGGCGAGCCCGAGCCAGCG 1236
D 1177 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 979 GCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 920
QY 1237 CTTCCCGCGGTCTTAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1296
D 1237 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 919 GGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 860
QY 1297 CTAAGACACCTGCTGTATCTCCACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1351
D 1297 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 859 CCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 805

RESULT 14
AG081217/c
LOCUS
DEFINITION
Pan troglodytes DNA, clone PTB-077J21.F, genomic survey sequence.
ACCESSION
AG081217
KEYWORDS
GSS.
SOURCE
Pan troglodytes
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Primates; Catarrhini; Hominidae; Pan.

```



```

REFERENCE
1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of library PTB
JOURNAL Unpublished
REFERENCE
2
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submissio3
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbese@gsc.riken.go.jp, URL:http://hgpc.gsc.riken.go.jp/).
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB. This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY
Vector : pSL145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES
source
1..900
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-077J21.F"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 66 a 304 c 394 g 34 t 102 others
ORIGIN
Query Match 3.1%; Score 48; DB 17; Length 900;
Best Local Similarity 49.0%; Pred. No.12;
Matches 96; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
Qy 1176 GTGCGCTGGTGGCGGAGCTGAGACTAGAGACTGCGGCGAGCGCCGAGCCGAGC 1235
Db 479 GGGCCCGGGNCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
Qy 1236 GTTCCCGCGCGCTTTAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
Db 419 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
Qy 1296 CCTAGACACCTGCTACCTCCACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC
Db 359 CCCCCCCCCNNCNCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCCNCC
Qy 1356 TTGTGCTGGCGGCGCT 1371
Db 299 TTATTGGAGGGGAT 284

RESULT 15
CNS010EW/3
LOCUS 1009 bp DNA linear GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN03P19 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL098882
VERSION AL098882.1 GI:5610493
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1009)
GENSCOPE
AUTHORS Direct Submission
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :

```

```

BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - . This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
peloBAC11.
FEATURES
source
1..1009
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN03P19"
/cclone_lib="DrosBAC"
/plasmid="pBeloBAC11"
/note="end : SP6"
BASE COUNT 147 a 377 c 178 g 64 t 243 others
ORIGIN
Query Match 3.1%; Score 47.6; DB 17; Length 1009;
Best Local Similarity 37.3%; Pred. No.14;
Matches 85; Conservative 50; Mismatches 88; Indels 5; Gaps 1;
Qy 1186 GTCGCGGGAACGTGACTAGAGACTGCGGCGAGCCCGAGCCCGAGCGCTCCCGCG 1245
Db 858 SCCNCGRGCGGGCGGACGCGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSG
Qy 1246 GGTCTTAGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
Db 798 SGGSSGGCGACGCGCGGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG
Qy 1306 CTGCTGTACCTCCACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG
Db 738 CSGTGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
Qy 1366 CGGCTGAACGTGCGCCCTTTAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
Db 683 GGGGKGGMGCGCGGGSGGGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGGSGGG
Search completed: February 4, 2003, 03:26:08
Job time : 2926.52 secs

```



```

: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: ptz9pt-F1s
: US-08-232-463-14

Query Match          2.7%  Score 40.6; DB 1; Length 7218;
Best Local Similarity 4.2%; pred.No. 0.34; Mismatches 125; Indels 0; Gaps 0;
Matches 13; Conservative 171;

QY 298 CTGCCCTCAAGTGGCTCAAAAGGATCTTTGTCAAGGTAGGAGACCTTGTGGCTCCA 357
      |||| | ||| ||| | : : : : : : : : : : : : : : : : : : : : : :
Db 1043 CTGCAGGTGAGGAGCTGGCATYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1102

QY 358 CGTGACCTCCAGGCGCTGTGGCTCTTCAAGGCTCTCTCGAGTCTCTTAAGAT 417
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1203 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1162

QY 418 CTCCTTCAGGCGAGATCATATATAGACTCTTCACAGTTTGACTGATTTTGGCCAGA 477
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1163 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1222

QY 478 ATAGGTGACATTTAGTTTGTGGCTTGATGAATGACATTAATATTAGACATATGGTGT 537
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1223 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1282

QY 538 GTAGGCTGCACTTCTACTCTTGCCITTTTGTGGCCCTCCAGTGTTCGGGTAGTTT 597
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1283 YVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 1342

QY 598 GCTCCCGCT 606
      : : : : : :
Db 1343 YVYVYVYV 1351

RESULT 3
US-08-483-533-4
: Sequence 4, Application US/08483533
: Patent No. 6172047
: GENERAL INFORMATION:
: APPLICANT: Roizman, Bernard
: APPLICANT: Chou, Joany
: TITLE OF INVENTION: Method for Treating Tumorigenic
: TITLE OF INVENTION: Diseases
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/483,533
: FILING DATE: 07-MAR-95
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/419,853
: FILING DATE: 11-APR-95
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/861,233
: FILING DATE: 31-MAR-92
: ATTORNEY/AGENT INFORMATION:
: NAME: Zeller, James P.
: REGISTRATION NUMBER: 28,491
: REFERENCE/DOCKET NUMBER: 28097/32742
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312/474-6300
: TELEFAX: 312/474-0448
: TELEX: 25-3856
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 595 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-483-533-4

Query Match          2.5%  Score 38.8; DB 4; Length 595;
Best Local Similarity 49.2%; Pred.No. 0.33; Mismatches 132; Indels 3; Gaps 1;
Matches 131; Conservative 0;

QY 1171 CCACTGTGTCCTTGGCTGCGGGAACGTGACATAGAGTCTGCGGCGCAGCCCGAGC 1230
      |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
Db 99 CCGCAGAGACACTGCGCGGCTGCGCTCGACGCGCGGGGAGGGGCGCGGAGC 158

QY 1231 CCAGCGGTCCCGCGGCTCTTAGCGCGGCGCGCGCGGCGGGAAGGGGACGACGACC 1290
      |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
Db 159 CCCC CGGAGACCCCGCGGACCCCGCGGACCCCGCGGACCCCGCGGACCCCGCGG 218

QY 1291 CGGACCCCTAAGACACTGCTGTACCTCCACCCCGCACCC---CACCCACCTCCCGCCAAAC 1347
      |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
Db 219 CGACCCCGGAGAACCCCGCGGACCCCGCGGACCCCGCGGCTGCTCTCGCCCGACG 278

QY 1348 TCCTAGATGTCTGTTGGGCGGTGACAGTCTGCCCGTTTAAGGGGCGGGGCGGCTCCA 1407
      ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
Db 279 TCCGGGTGCGCCACTGTGTGCTGTGGCTCTGCGCGCGCGGCTGCGCGCGGCTCGT 338

QY 1408 GTGCTTCTCTGAGTGAAGTAACT 1433
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 339 GGGCCCGGAGCGGGCGGACCGCGGT 364

RESULT 4
US-09-283-471A-4
: Sequence 4, Application US/09283471A
: Patent No. 6346673
: GENERAL INFORMATION:
: APPLICANT: Roizman, Bernard
: APPLICANT: Chou, Joany
: TITLE OF INVENTION: Method For Treating Tumorigenic Diseases
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower, 233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
: COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/283,471A
FILING DATE: 04-APR-1999
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/861,233
FILING DATE: 31-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/419,853
FILING DATE: 11-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/483,533
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, James P.
REGISTRATION NUMBER: 28,491
REFERENCE/DOCKET NUMBER: 27373/32742A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 595 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-283-471A-4

Query Match 2.5%; Score 38.8; DB 4; Length 595;

Best Local Similarity 49.28; Pred. No. 0.33;

Matches 131; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

QY 1171 CCATGTGTCCTTGGTGGCGGAACTGGACTAGAGAGTCTGGCGGCGAGCCCGGAGC 1230
DB 99 CCGAGAGACACTGGCGGCTGTGGCTTGGACGCGCGGGGGAGGGGGCGCGGAGC 158
QY 1231 CCAGCGCTTCCCGGCGCTTAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1290
DB 159 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 218
QY 1291 CGGACCTTAAAGACACTGCTGTACCTTCCACCCCAACCC---CACCCACCTCCCGCCCAAC 1347
DB 215 CGACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 278
QY 1348 TCCCTAGATGTGTGTGGCGGCTGAAGTGTGCGGCTTAAAGGCGGCGGCGGCGGCTCCA 1407
DB 279 TCGGGGTGCGGCACTGTGTGTGGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCTCGT 338
QY 1408 CGTGTCTTCTGCTGAGTGAAGT 1433
DB 339 GGGCGCGGAGCGGCGGCGGCGGCT 364

RESULT 5

US-08-483-533-36

Sequence 36, Application US/08483533

Patent No. 6172047

GENERAL INFORMATION:

APPLICANT: Roizman, Bernard

APPLICANT: Chou, Joany

TITLE OF INVENTION: Method for Treating Tumorigenic

TITLE OF INVENTION: Diseases

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Born

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,533
FILING DATE: 07-MAR-95
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/419,853
FILING DATE: 11-APR-95
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/861,233
FILING DATE: 31-MAR-92
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, James P.
REGISTRATION NUMBER: 28,491
REFERENCE/DOCKET NUMBER: 28097/32742
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 1327 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-483-533-36

Query Match 2.5%; Score 38.8; DB 4; Length 1327;

Best Local Similarity 49.28; Pred. No. 0.48;

Matches 131; Conservative 0; Mismatches 132; Indels 3; Gaps 1;

QY 1171 CCATGTGTCCTTGGTGGCGGAACTGGACTAGAGAGTCTGGCGGCGAGCCCGGAGC 1230
DB 540 CCGAGAGACACTGGCGGCTGTGGCTTGGACGCGCGGGGGAGGGGGCGCGGAGC 599
QY 1231 CCAGCGCTTCCCGGCGCTTAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1290
DB 600 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 659
QY 1291 CGGACCTTAAAGACACTGCTGTACCTTCCACCCCAACCC---CACCCACCTCCCGCCCAAC 1347
DB 660 CGACCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 719
QY 1348 TCCCTAGATGTGTGTGGCGGCTGAAGTGTGCGGCTTAAAGGCGGCGGCGGCGGCTCCA 1407
DB 720 TCGGGGTGCGGCACTGTGTGTGGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCTCGT 779
QY 1408 CGTGTCTTCTGCTGAGTGAAGT 1433
DB 780 GGGCGCGGAGCGGCGGCGGCGGCT 805

RESULT 6

US-09-283-471A-36

Sequence 36, Application US/09283471A

Patent No. 6340873

GENERAL INFORMATION:

APPLICANT: Roizman, Bernard

APPLICANT: Chou, Joany

TITLE OF INVENTION: Method For Treating Tumorigenic Diseases

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Born

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

1	EARLIER APPLICATION NUMBER: 60/038,621	1	EARLIER FILING DATE: 1997-04-11
2	EARLIER FILING DATE: 1997-03-07	2	EARLIER APPLICATION NUMBER: 60/043,677
3	EARLIER APPLICATION NUMBER: 60/040,626	3	EARLIER FILING DATE: 1997-04-11
4	EARLIER FILING DATE: 1997-03-07	4	EARLIER APPLICATION NUMBER: 60/043,315
5	EARLIER APPLICATION NUMBER: 60/040,334	5	EARLIER FILING DATE: 1997-04-11
6	EARLIER FILING DATE: 1997-03-07	6	EARLIER APPLICATION NUMBER: 60/048,974
7	EARLIER APPLICATION NUMBER: 60/040,336	7	EARLIER FILING DATE: 1997-06-06
8	EARLIER FILING DATE: 1997-03-07	8	EARLIER APPLICATION NUMBER: 60/056,886
9	EARLIER APPLICATION NUMBER: 60/040,163	9	EARLIER FILING DATE: 1997-08-22
10	EARLIER FILING DATE: 1997-03-07	10	EARLIER APPLICATION NUMBER: 60/056,977
11	EARLIER APPLICATION NUMBER: 60/047,600	11	EARLIER FILING DATE: 1997-08-22
12	EARLIER FILING DATE: 1997-05-23	12	EARLIER APPLICATION NUMBER: 60/056,889
13	EARLIER APPLICATION NUMBER: 60/047,615	13	EARLIER FILING DATE: 1997-08-22
14	EARLIER FILING DATE: 1997-05-23	14	EARLIER APPLICATION NUMBER: 60/056,893
15	EARLIER APPLICATION NUMBER: 60/047,597	15	EARLIER FILING DATE: 1997-08-22
16	EARLIER FILING DATE: 1997-05-23	16	EARLIER APPLICATION NUMBER: 60/056,630
17	EARLIER APPLICATION NUMBER: 60/047,502	17	EARLIER FILING DATE: 1997-08-22
18	EARLIER FILING DATE: 1997-05-23	18	EARLIER APPLICATION NUMBER: 60/056,878
19	EARLIER APPLICATION NUMBER: 60/047,633	19	EARLIER FILING DATE: 1997-08-22
20	EARLIER FILING DATE: 1997-05-23	20	EARLIER APPLICATION NUMBER: 60/056,662
21	EARLIER APPLICATION NUMBER: 60/047,583	21	EARLIER FILING DATE: 1997-08-22
22	EARLIER FILING DATE: 1997-05-23	22	EARLIER APPLICATION NUMBER: 60/056,872
23	EARLIER APPLICATION NUMBER: 60/047,617	23	EARLIER FILING DATE: 1997-08-22
24	EARLIER FILING DATE: 1997-05-23	24	EARLIER APPLICATION NUMBER: 60/056,882
25	EARLIER APPLICATION NUMBER: 60/047,618	25	EARLIER FILING DATE: 1997-08-22
26	EARLIER FILING DATE: 1997-05-23	26	EARLIER APPLICATION NUMBER: 60/056,637
27	EARLIER APPLICATION NUMBER: 60/047,503	27	EARLIER FILING DATE: 1997-08-22
28	EARLIER FILING DATE: 1997-05-23	28	EARLIER APPLICATION NUMBER: 60/056,903
29	EARLIER APPLICATION NUMBER: 60/047,592	29	EARLIER FILING DATE: 1997-08-22
30	EARLIER FILING DATE: 1997-05-23	30	EARLIER APPLICATION NUMBER: 60/056,888
31	EARLIER APPLICATION NUMBER: 60/047,581	31	EARLIER FILING DATE: 1997-08-22
32	EARLIER FILING DATE: 1997-05-23	32	EARLIER APPLICATION NUMBER: 60/056,879
33	EARLIER APPLICATION NUMBER: 60/047,584	33	EARLIER FILING DATE: 1997-08-22
34	EARLIER FILING DATE: 1997-05-23	34	EARLIER APPLICATION NUMBER: 60/056,880
35	EARLIER APPLICATION NUMBER: 60/047,500	35	EARLIER FILING DATE: 1997-08-22
36	EARLIER FILING DATE: 1997-05-23	36	EARLIER APPLICATION NUMBER: 60/056,894
37	EARLIER APPLICATION NUMBER: 60/047,587	37	EARLIER FILING DATE: 1997-08-22
38	EARLIER FILING DATE: 1997-05-23	38	EARLIER APPLICATION NUMBER: 60/056,911
39	EARLIER APPLICATION NUMBER: 60/047,492	39	EARLIER FILING DATE: 1997-08-22
40	EARLIER FILING DATE: 1997-05-23	40	EARLIER APPLICATION NUMBER: 60/056,636
41	EARLIER APPLICATION NUMBER: 60/047,598	41	EARLIER FILING DATE: 1997-08-22
42	EARLIER FILING DATE: 1997-05-23	42	EARLIER APPLICATION NUMBER: 60/056,874
43	EARLIER APPLICATION NUMBER: 60/047,613	43	EARLIER FILING DATE: 1997-08-22
44	EARLIER FILING DATE: 1997-05-23	44	EARLIER APPLICATION NUMBER: 60/056,910
45	EARLIER APPLICATION NUMBER: 60/047,582	45	EARLIER FILING DATE: 1997-08-22
46	EARLIER FILING DATE: 1997-05-23	46	EARLIER APPLICATION NUMBER: 60/056,864
47	EARLIER APPLICATION NUMBER: 60/047,596	47	EARLIER FILING DATE: 1997-08-22
48	EARLIER FILING DATE: 1997-05-23	48	EARLIER APPLICATION NUMBER: 60/056,631
49	EARLIER APPLICATION NUMBER: 60/047,612	49	EARLIER FILING DATE: 1997-08-22
50	EARLIER FILING DATE: 1997-05-23	50	EARLIER APPLICATION NUMBER: 60/056,845
51	EARLIER APPLICATION NUMBER: 60/047,632	51	EARLIER FILING DATE: 1997-08-22
52	EARLIER FILING DATE: 1997-05-23	52	EARLIER APPLICATION NUMBER: 60/056,892
53	EARLIER APPLICATION NUMBER: 60/047,601	53	EARLIER FILING DATE: 1997-08-22
54	EARLIER FILING DATE: 1997-05-23	54	EARLIER APPLICATION NUMBER: 60/057,761
55	EARLIER APPLICATION NUMBER: 60/043,580	55	EARLIER FILING DATE: 1997-08-22
56	EARLIER FILING DATE: 1997-04-11	56	EARLIER APPLICATION NUMBER: 60/047,595
57	EARLIER APPLICATION NUMBER: 60/043,568	57	EARLIER FILING DATE: 1997-05-23
58	EARLIER FILING DATE: 1997-04-11	58	EARLIER APPLICATION NUMBER: 60/047,599
59	EARLIER APPLICATION NUMBER: 60/043,314	59	EARLIER FILING DATE: 1997-05-23
60	EARLIER FILING DATE: 1997-04-11	60	EARLIER APPLICATION NUMBER: 60/047,588
61	EARLIER APPLICATION NUMBER: 60/043,569	61	EARLIER FILING DATE: 1997-05-23
62	EARLIER FILING DATE: 1997-04-11	62	EARLIER APPLICATION NUMBER: 60/047,585
63	EARLIER APPLICATION NUMBER: 60/043,311	63	EARLIER FILING DATE: 1997-05-23
64	EARLIER FILING DATE: 1997-04-11	64	EARLIER APPLICATION NUMBER: 60/047,586
65	EARLIER APPLICATION NUMBER: 60/043,671	65	EARLIER FILING DATE: 1997-05-23
66	EARLIER FILING DATE		


```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4481 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 163...4099
; US-09-041-886-18

Query Match 2.5%; Score 38.6; DB 4; Length 4481;
Best Local Similarity 52.1%; Pred. No. 0.95;
Matches 86; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Qy 1187 CTGCGGGAGCTGACTAGAGAGTCTCGCGCGAGCCCGGAGCCCGCGCTTCCCGGCG 1246
Db 194 CCGGGTGGCCACCGAGTCTCGCGCGCTTCCCGCGAGCGTGGCGGGGTGGCGCTCGC 253
Qy 1247 GCTTAGCGCGCGCGCGCGCGGAGGAGGAGCGAGACCGCGGACCCCTAAGACACC 1306
Db 254 TCCAGCGCGCGCGCGCGGAGCGGCGGCGGCGGTGGCGCGCGCGCGCGCGGTATC 313
Qy 1307 TCGTGTACCTTCCAGCCCGCCAGCCCGACCCAGCCCGTCCCGCAATCC 1351
Db 314 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCC 358

RESULT 11
US-08-709-838-1
; Sequence 1, Application US/08709838
; Patent No. 6140064
; GENERAL INFORMATION:
; APPLICANT: Loetscher, Marcel
; APPLICANT: Moser, Bernhard
; TITLE OF INVENTION: IP-10/MIG RECEPTOR DESIGNATED CXCR3,
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0. Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 31-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/709,838
; FILING DATE: 10-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook Esq., David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: TK196-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1670 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 69...1172
; US-08-829-839-1

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1670 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 69...1172
; US-08-709-838-1

Query Match 2.5%; Score 38.2; DB 3; Length 1670;
Best Local Similarity 59.8%; Pred. No. 0.78;
Matches 64; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 714 TCCTGTGTTTATCAGAGGAGCGCTCATATATGAATTAAGAGGGGGTGGTCC 773
Db 1562 TCTTTATTTTATGTCTAAATCCIGCTTAAACTTTCATTAACAAGATCGTCAGGA 1621
Qy 774 CATATTGTCGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 820
Db 1622 CTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 1668

RESULT 12
US-08-829-839-1
; Sequence 1, Application US/08829839
; Patent No. 6184358
; GENERAL INFORMATION:
; APPLICANT: Loetscher, Marcel
; APPLICANT: Moser, Bernhard
; APPLICANT: Qin, Shixin
; APPLICANT: Mackay, Charles R.
; TITLE OF INVENTION: IP-10/MIG RECEPTOR DESIGNATED CXCR3,
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0. Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 31-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/709,838
; FILING DATE: 10-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Brook Esq., David E.
; REGISTRATION NUMBER: 22,592
; REFERENCE/DOCKET NUMBER: TK196-01A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1670 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 69...1172
; US-08-829-839-1
```


FEATURE:
NAME/KEY: allele
LOCATION: 93590..93736
OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
FEATURE:
NAME/KEY: allele
LOCATION: 97099..97145
OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
FEATURE:
NAME/KEY: allele
LOCATION: 97130..97177
OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
FEATURE:
NAME/KEY: allele
LOCATION: 99075..99121
OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
FEATURE:
NAME/KEY: allele
LOCATION: 99094..99140
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
FEATURE:
NAME/KEY: allele
LOCATION: 103783..103828
OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106965
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
FEATURE:
NAME/KEY: allele
LOCATION: 106918..106965
OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
FEATURE:
NAME/KEY: allele
LOCATION: 108054..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
FEATURE:
NAME/KEY: allele
LOCATION: 108054..108130
OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
FEATURE:
NAME/KEY: allele
LOCATION: 108127..108177
OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
FEATURE:

Query Match 2.5%; Score 37.6; DB 4; Length 162450;
Best Local Similarity 52.2%; Pred. No. B.4;
Matches 84; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 686 TCACAAATGATACAAACTAAATACAAAGTCTGTGTATATATACAGGAGCGTGATCAA 745
DB 109593 TAAACATTTTTTACTATATATATATATAGCCCAATTTTATTTCCCTCCCTGTTCTG 109534
QY 746 TATAATGAATTAAGAGGGGCTGTCCCATATATGTTCTGTGTTTGTGTTTGTGTTT 805
DB 109533 CTGCATACTGIGGCCAAATGTCAGTACGCTAAATGATTTTTTTTTTTTTTTTTTT 109474
QY 806 TTTTGTGTTTGTGCGCCCTCTCTCTCAATTAATGAAG 846
DB 109473 TTTTGTGTTTGTGAGACAGAGTCTCATCGATCCCAAG 109433
RESULT 14
US-08-018-977C-4/c
; Sequence 4, Application JS/08018977C
; Patent No. 56866C1
; GENERAL INFORMATION:
; APPLICANT: Weber, Peter C.
; TITLE OF INVENTION: DNA Sequences Encoding Mutant Antiviral
; TITLE OF INVENTION: Regulatory Proteins
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: The Pennsylvania State University
; STREET: 113 Technology Center
; CITY: University Park
; STATE: PA
; COUNTRY: US
; ZIP: 16802-7000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS 6.22
; SOFTWARE: Microsoft Word for Windows 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/018,977C
; FILING DATE: 18-FEB-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/726,671
; FILING DATE: 05-JUL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas J. Monahan
; REGISTRATION NUMBER: 29,835
; REFERENCE/DOCKET NUMBER: 9-1039/2
; TELEPHONE: (814) 865-6277
; TELEFAX: (814) 865-3591
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 936 bases
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-018-977C-4

Query Match 2.5%; Score 37.6; DB 1; Length 936;
Best Local Similarity 54.3%; Pred. No. 0.87;
Matches 76; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 1212 CTGGGGCGCAGCCCGGAGCGCTTCCCGCGGCTTAGAGCGGGGGCGCGGGGG 1271
DB 785 CTCGAGCGCAGCGCGCTCTCTCTGCTGACAGAGGCGGGGAGCGGGGCGCG 726
QY 1272 GGAAGGGGACGACAGCGCGGACCGCTTAAGACACCTGTACCTCCACCCACCCA 1331
DB 725 GAGGGGGGCGCGCGGAGGGCGCGTCCACCTCCACCGCGCGCGCGCGCGCG 666
QY 1332 CCCAGCTCCCGCAACTCC 1351
DB 665 CGCGCGCAGCTCGCACGCGC 646

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 12:46:48 ; Search time 154.036 Seconds
(without alignments)
8245.650 Million cell updates/sec

Title: US-09-596-141C-3_COPY_1080_1643

Perfect score: 564

Sequence: 1 gaggctctctcccccaatc.....caggggcaggggggagctc 564

Scoring table: IDENTITY-NUC

Gapop 10.0 ; Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*

2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*

3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*

4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*

5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*

6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*

7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*

8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*

9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*

10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*

11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*

12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*

13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*

14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*

15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*

16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*

17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*

18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*

19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*

20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*

21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*

22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*

23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*

24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	564	100.0	1643	22	AAF24681
2	564	100.0	1643	22	AAF24703
3	552	97.9	1197	24	ABL58400
4	548	97.2	183999	22	AAF92831
5	547.4	97.1	3231	24	AAF24265
6	473.4	83.9	2910	24	AAD37267
7	119.4	21.2	227	21	AA096515
8	98	17.4	7260	22	AAD21326
9	98	17.4	7260	22	AA170315

10	91	16.1	221	24	AAD37268
11	83.4	15.9	736	22	AAE07432
12	83.4	15.9	1556	22	AAH18606
13	89	15.8	763	22	AAH04729
14	89	15.8	1750	22	AAH17451
15	77.2	13.7	7281	22	AAK51683
16	75.4	13.4	7086	22	ABA09200
17	75.4	13.4	7086	22	AAK52667
18	75	13.3	9854	22	AA06121
19	67	11.9	10442	22	AAF24580
20	67	11.9	10442	22	AAF24702
21	60	10.6	10474	22	AAF24685
22	60	10.6	10474	22	AAF24686
23	60	10.6	10474	22	AAF24707
24	60	10.6	10474	22	AAF24708
25	45	8.0	114955	20	AAK53491
26	38.8	6.9	595	14	AAQ50041
27	38.8	6.9	1335	13	AAQ23295
28	38.6	6.8	4367	19	AAV3027C
29	38.6	6.8	4481	19	AAV06552
30	38.6	6.8	4481	20	AAZ23428
31	38.6	6.8	18660	21	AAA58472
32	37.6	6.7	381	24	ASN73567
33	37.6	6.7	936	18	AAT95102
34	37.4	6.6	1599	16	AAQ97701
35	37.4	6.6	12042	23	AAK76999
36	37.4	6.6	175737	24	ABK83571
37	37	6.6	1555	21	AAA64196
38	36.8	6.5	565	8	AAW70930
39	36.8	6.5	565	22	AAF32043
40	36.8	6.5	19408	22	AAK42003
41	36.8	6.5	19408	22	AAK87230
42	36.8	6.5	19408	22	AAK90644
43	36.6	6.5	128139	24	AA164291
44	36.6	6.5	133719	21	AAK64754
45	36.4	6.5	38	22	AAF93064

ALIGNMENTS

RESULT: 1
AAF24681
ID AAF24681 standard; DNA; 1643 bp.

XX AAF24681;

XX AC

XX 20-APR-2001 (first entry)

XX DT Nucleotide sequence of the 5' flanking region of the human ABC1 gene.

XX DE Human; adenosine triphosphate binding cassette protein 1; ABC1;

XX KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;

XX KW chromosome 9q22-q31; heart disease; hypercholesterolemia;

XX KW atherosclerosis; cholesterol transport; ss.

XX OS Homo sapiens.

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

XX XX

```
XX Adenosine triphosphate (ATP) binding cassette (ABC) polynucleotide,
PI useful for the development of agents for the treatment of heart disease
PI and other disorders associated with hypercholesterolemia and
PT atherosclerosis -
XX
PS Claim 1: Page 143-144; 215pp; English.
XX
CC The present sequence represents the 5' flanking region of the human
CC adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABC1
CC resides in cell membranes and utilises ATP hydrolysis to transport a wide
CC variety of substrates across the plasma membrane. ABC1 is a pivotal
CC protein in the apolipoprotein-mediated mobilisation of intracellular
CC cholesterol stores. ABC1 is defective in Tangier disease, a genetic
CC disorder characterised by abnormal HDL-cholesterol metabolism. The ABC1
CC gene is localised to chromosome 9q22-9q31. The ABC1 genes and proteins
CC are useful for developing pharmaceutical agents for the treatment of
CC heart disease and other disorders associated with hypercholesterolemia
CC and atherosclerosis. The genes are useful for developing screening assays
CC to screen for compounds that regulate the expression of genes associated
CC with cholesterol transport. The genes and proteins are also useful for
CC other disorders associated with hypercholesterolemia.
XX
SQ Sequence 1643 BP; 370 A; 413 C; 457 G; 403 T; 0 other;
Query Match 100.0%; Score 564; DB 22; Length 1643;
Best Local Similarity 100.0%; Pred. No. 3.9e-140;
Matches 564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGCTCTCTCTCCCAATCCCTCCCGCTGAGGAACTAACAAAGGAAAAAAT 60
DB 1080 GAGCTCTCTCTCCCAATCCCTCCCGCTGAGGAACTAACAAAGGAAAAAAT 1139
QY 61 TCGGGAAGCAGGATTAGAGGAAGCAAAATCCACTGGTGCCTTGGCTCCCGGAACGT 120
DB 1140 TCGGGAAGCAGGATTAGAGGAAGCAAAATCCACTGGTGCCTTGGCTCCCGGAACGT 1199
QY 121 GGACTAGAGAGTCTGCGGCGCAGCCCGAGCCCGAGCCCGAGCCGCTTCTAGCGCGG 180
DB 1200 GGACTAGAGAGTCTGCGGCGCAGCCCGAGCCCGAGCCCGAGCCGCTTCTAGCGCGG 1259
QY 181 GGGCCCGGGGGGGAAGGGGAGCGAGACCGGAGACCCCTAAGACACCTGTGTACCCCTC 240
DB 1260 GGGCCCGGGGGGGAAGGGGAGCGAGACCGGAGACCCCTAAGACACCTGTGTACCCCTC 1319
QY 241 ACCCCACCCACCCACCTCCCTCCCACTCCCTAGATGTGTGCTGGCGGCTGACGCTG 300
DB 1320 ACCCCACCCACCCACCTCCCTCCCACTCCCTAGATGTGTGCTGGCGGCTGACGCTG 1379
QY 301 CCGGTTAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 360
DB 1380 CCGGTTAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1439
QY 361 ACAGAGCCCGGGAACGGGGGGGGGAGGAGGAGGAGACAGCGCTTGTACCGATAGTAACT 420
DB 1440 ACAGAGCCCGGGAACGGGGGGGGGAGGAGGAGGAGACAGCGCTTGTACCGATAGTAACT 1499
QY 421 CTGGCTCGGTGAGCGCGGAACTATATAAAGAACTATCTCCCGCAAAACCCGTAATG 480
DB 1500 CTGGCTCGGTGAGCGCGGAACTATATAAAGAACTATCTCCCGCAAAACCCGTAATG 1559
QY 481 CGAGCAGAGTGTAGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 540
DB 1560 CGAGCAGAGTGTAGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 619
QY 541 GCGGCAGGGGAGGCGGGGAGGCTC 564
DB 1620 GCGGCAGGGGAGGCGGGAGGCTC 1643
RESULT 2
AAF24703
```

```
ID AAF24703 standard; DNA; 1643 BP.
XX
AC AAF24703.
XX
DT 20-APR-2001 (first entry)
XX
DE Nucleotide sequence of the 5' flanking region of the human ABC1 gene.
XX
KW Human: adenosine triphosphate binding cassette protein 1; ABC1;
KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
KW atherosclerosis; cholesterol transport; ss.
XX
OS Homo sapiens.
XX
PN W0200078971-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000W0-US16591.
XX
PR 18-JUN-1999; 990S-0140264.
PR 14-SEP-1999; 990S-0153672.
PR 19-NOV-1999; 990S-0166573.
XX
CV (CVTH-) CV THERAPEUTICS INC.
XX (UNIW) UNIV WASHINGTON.
XX
Lawn RM, Wade D, Oram JF, Garvin M;
XX
WPI; 2001-137811/14.
XX
PT Adenosine triphosphate (ATP) binding cassette protein (ABC) 1
PT polynucleotides and polypeptides, useful for treatment of heart disease
PT and other disorders associated with hypercholesterolemia and
PT atherosclerosis -
XX
PS Disclosure; Page 138-139; 211pp; English.
XX
CC The present sequence represents the 5' flanking region of the human
CC adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABC1
CC resides in cell membranes and utilises ATP hydrolysis to transport a wide
CC variety of substrates across the plasma membrane. ABC1 is a pivotal
CC protein in the apolipoprotein-mediated mobilisation of intracellular
CC cholesterol stores. ABC1 is defective in Tangier disease, a genetic
CC disorder characterised by abnormal HDL-cholesterol metabolism. The ABC1
CC gene is localised to chromosome 9q22-9q31. The ABC1 genes and proteins
CC are useful for developing pharmaceutical agents for the treatment of
CC heart disease and other disorders associated with hypercholesterolemia
CC and atherosclerosis. The genes are useful for developing screening assays
CC to screen for compounds that regulate the expression of genes associated
CC with cholesterol transport. The genes and proteins are also useful for
CC other disorders associated with hypercholesterolemia.
XX
SQ Sequence 1643 BP; 370 A; 413 C; 457 G; 403 T; 0 other;
```

```
Query Match 100.0%; Score 564; DB 22; Length 1643;
Best Local Similarity 100.0%; Pred. No. 3.9e-140;
Matches 564; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAGCTCTCTCTCCCAATCCCTCCCGCTGAGGAACTAACAAAGGAAAAAAT 60
DB 1080 GAGCTCTCTCTCCCAATCCCTCCCGCTGAGGAACTAACAAAGGAAAAAAT 1139
QY 61 TCGGGAAGCAGGATTAGAGGAAGCAAAATCCACTGGTGCCTTGGCTCCCGGAACGT 120
DB 1140 TCGGGAAGCAGGATTAGAGGAAGCAAAATCCACTGGTGCCTTGGCTCCCGGAACGT 1199
QY 121 GGACTAGAGAGTCTGCGGCGCAGCCCGAGCCCGAGCCCGAGCCGCTTCTAGCGCGG 180
DB 1200 GGACTAGAGAGTCTGCGGCGCAGCCCGAGCCCGAGCCCGAGCCGCTTCTAGCGCGG 1259
```


XX 01-SEP-2000; 2000WO-1B01492.
XX
XX 01-SEP-1999; 99US-0151977.
XX 15-MAR-2000; 2000US-3526193.
XX 23-JUN-2000; 2000US-0213958.
XX (UYBR-) UNIV BRITISH COLUMBIA.
XX (XENO-) XENON GENETICS INC.
XX
XX Hayden MR, Brooks-Wilson AR, Pimstone SN, Clee SM;
XX WPI; 2001-244356/25.
XX
XX Treating a lower than normal high density lipoprotein-cholesterol
XX (HDL-C) level, a higher than normal triglyceride level, or a
XX cardiovascular disease, by administering a compound that modulates LXR-
XX or RXR-mediated transcriptional activity -
XX
XX Claim 8; Fig 1; 317pp; English.
XX
XX The present invention relates to a method for treating a patient
XX diagnosed as having a lower than normal high density
XX lipoprotein-cholesterol (HDL-C) level, a higher than normal
XX triglyceride level, or a cardiovascular disease, involving
XX administering a compound that modulates LXR- or RXR-mediated
XX transcriptional activity or ABC1 expression or activity.
XX The LXR gene product may be used in an assay to identify
XX compounds useful for the treatment of a disease or condition selected a
XX lower than normal HDL cholesterol level, a higher than normal
XX triglyceride level, and a cardiovascular disease.
XX
XX Sequence 183999 BP; 49549 A; 37944 C; 41170 G; 54950 T; 386 other:
SQ
Query Match 97.2%; Score 548; DB 22; Length 183999;
Best Local Similarity 98.9%; Pred. No. 2.8e-135;
Matches 563; Conservative 0; Mismatches 1; Indels 5; Gaps 1:
QY 1 GAGCTCTCTCTCCCAATCCCTCCCTCCGCTGGAGAACTAACAAAGGAAAAAATAAT 60
DB 28312 GAGCTCTCTCTCCCAATCCCTCCCTCCGCTGGAGAACTAACAAAGGAAAAAATAAT 28371
QY 51 TCGGAAAGCAGGATTAGAGGAAGCAATTCACCTGGTCCCTGGCTGGCGGAAAGCT 120
DB 26372 TCGGAAAGCAGGATTAGAGGAAGCAATTCACCTGGTCCCTGGCTGGCGGAAAGCT 28431
QY 121 GGACTAGAGAGTCTCGGGGAGCCCGAGCCACGCTTCCCGCGGCTCTTAGCCCGGC 180
DB 28432 GGACTAGAGAGTCTCGGGGAGCCCGAGCCACGCTTCCCGCGGCTCTTAGCCCGGC 28491
QY 181 GGGCCCGGGGGGAAAGGGGAGCGAGCCGGGACCTTAAGACACCTGCTACCTCC 240
DB 28492 GGGCCCGGGGGGAAAGGGGAGCGAGCCGGGACCTTAAGACACCTGCTACCTCC 28551
QY 241 A-----CCCCACCCACCCACCTCCGCCCACTCCCTAGATGTGCTGGCGGCTGAA 295
DB 28552 ANNNNNCCCCACCCACCCACCTCCGCCCACTCCCTAGATGTGCTGGCGGCTGAA 28611
QY 296 CGTCCGCCCTTTAAGGGGCGCGCGCTCCACGTCCTTTCTGCTAGTGAAGTGAAGTA 355
DB 28612 CGTCCGCCCTTTAAGGGGCGCGCGCTCCACGTCCTTTCTGCTAGTGAAGTGAAGTA 28671
QY 356 CATAACAGAGCGCGGAAAGGGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 415
DB 28672 CATAACAGAGCGCGGAAAGGGGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 28731
QY 416 AAGCTCTGGCTGGTTCAGCCGAATCTATAAAGGAATAGTCCCGGCAAAACCCCGT 475
DB 28732 AAGCTCTGGCTGGTTCAGCCGAATCTATAAAGGAATAGTCCCGGCAAAACCCCGT 28791
QY 476 AATTGGAGGAGGAGTGTGGGGCGGGGACCCGCGAGACCCGAGCCCTCTCTCC 535
DB 28792 AATTGGAGGAGGAGTGTGGGGCGGGGACCCGCGAGACCCGAGCCCTCTCTCC 28851

QY 536 GGGCTGGCGCAGGCGAGGCGGGGAGCTC 564
DB 28852 GGGCTGGCGCAGGCGAGGCGGGGAGCTC 28880
RESULT 5
AAD37265
ID AAD37265 standard; DNA; 3231 3P.
XX
XX AAD37265;
XX
XX 21-AUG-2002 (first entry)
XX
XX Human ABC1 transcription regulatory DNA #1.
XX
XX Human; ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
XX cholesterol metabolism; hypercholesterolaemia; antisense therapy; GS.
XX
XX Homo sapiens.
XX
XX W0200183746-A2.
XX
XX 08-NOV-2001.
XX
XX 02-MAY-2001; 2001WO-EP05488.
XX
XX 02-MAY-2000; 2000US-201280P.
XX
XX (AVET) AVENTIS PHARMA SA.
XX
XX Rosier-Montus M, Prades C, Lemoine C, Naudin L, Deneffe P;
XX Brewer B, Duverger N, Remaley A, Santamarina-Fojo S;
XX
XX WPI: 2002-154404/20.
XX
XX Isolated nucleic acid useful for modifying the ATP-binding cassette 1
XX (ABC1) and screening for candidate modulatory compounds or substances
XX
XX Claim 1; Page 130-131; 152pp; English.
XX
XX The invention relates a nucleic acid which is capable of regulating the
XX transcription of human ATP-binding cassette 1 (ABC1) gene, which is a
XX causal gene for pathologies linked to a dysfunctioning of cholesterol
XX metabolism, including diseases such as atherosclerosis. Polynucleotides
XX of the invention are used to screen candidate molecules or substances
XX that are capable of modulating the transcription of the ABC1 gene. They
XX are used in antisense therapy. Compositions comprising sequences of the
XX invention are used to treat hypercholesterolaemia and atherosclerosis.
XX The present sequence is human ABC1 transcription regulating DNA.
SQ
Sequence 3231 BP; 809 A; 773 C; 876 G; 773 T; 0 other:

Query Match 97.1%; Score 547.4; DB 24; Length 3231;
Best Local Similarity 98.9%; Pred. No. 1.2e-135;
Matches 563; Conservative 0; Mismatches 1; Indels 5; Gaps 1:
QY 1 GAGCTCTCTCTCTCCCAATCCCTCCCTCCGCTGGAGAACTAACAAAGGAAAAAATAAT 60
DB 2415 GAGCTCTCTCTCTCCCAATCCCTCCCTCCGCTGGAGAACTAACAAAGGAAAAAATAAT 2475
QY 61 TCGGAAAGCAGGATTAGAGGAAGCAATTCACCTGGTGGCTTGGCTGGCGGAAAGCT 120
DB 2476 TCGGAAAGCAGGATTAGAGGAAGCAATTCACCTGGTGGCTTGGCTGGCGGAAAGCT 2535
QY 121 GGACTAGAGAGTCTCGGGGCGAGCCCGAGCCGCTTCCCGCGGCTCTTAGCCCGGC 180
DB 2536 GGACTAGAGAGTCTCGGGGCGAGCCCGAGCCGCTTCCCGCGGCTCTTAGCCCGGC 2595
QY 181 GGGCCCGGGGGGAAAGGGGAGCGAGCCGGGACCTTAAGACACCTGCTGTACCTCC 240
DB 2596 GGGCCCGGGGGGAAAGGGGAGCGAGCCGGGACCTTAAGACACCTGCTGTACCTCC 2655


```

FT protein_bind /bound_moiety= "PPAR/NKX2.5/PPAR"
FT 2051..2059 /*tag= ak
FT /bound_moiety= "GATA"
FT 2104..2111 /*tag= al
FT /bound_moiety= "SOX5"
FT 2114..2152 /*tag= am
FT /bound_moiety= "SYR/HFH/HNF3beta"
FT 2221..2228 /*tag= an
FT /bound_moiety= "MZF1"
FT 2234..2248 /*tag= ao
FT /bound_moiety= "IK2/NFKappaB/CREL"
FT 2259..2272 /*tag= ap
FT /bound_moiety= "LMO2COM/GATA"
FT 2289..2306 /*tag= aq
FT /bound_moiety= "MZF1/SRY"
FT 2313..2318 /*tag= ar
FT /bound_moiety= "PPAR"
FT 2321..2326 /*tag= as
FT /bound_moiety= "PPAR"
FT 2335..2342 /*tag= at
FT /bound_moiety= "MZF1"
FT 2361..2384 /*tag= au
FT /bound_moiety= "HNF3beta/SRY/EVIL"
FT 2426..2433 /*tag= av
FT /bound_moiety= "MZF1"
FT 2442..2451 /*tag= aw
FT /bound_moiety= "AP4"
FT 2455..2466 /*tag= ax
FT /bound_moiety= "SRY"
FT 2491..2498 /*tag= ay
FT /bound_moiety= "STAT"
FT 2524..2534 /*tag= az
FT /bound_moiety= "STAT/PPAR"
FT 2536..2541 /*tag= ba
FT /bound_moiety= "PPAR"
FT 2589..2600 /*tag= bb
FT /bound_moiety= "AP2"
FT 2610..2617 /*tag= bc
FT /bound_moiety= "MZF1"
FT 2634..2648 /*tag= bd
FT /bound_moiety= "LMO2COM/MYC/E47"
FT 2657..2672 /*tag= be
FT /bound_moiety= "RRE31"
FT 2680..2698 /*tag= bf
FT /bound_moiety= "MZF1/CMYB"
FT 2728..2740 /*tag= bg
FT /bound_moiety= "SP1/GC"
FT 2743..2757 /*tag= bh
FT /bound_moiety= "USF/NMYC/ARNT"

```

```

FT protein_bind 2758..2773 /*tag= bi
FT /bound_moiety= "NFE2A1"
FT 2774..2787 /*tag= bj
FT /bound_moiety= "XFD1/HFH"
FT 2794..2806 /*tag= bk
FT /bound_moiety= "GC/SP1/MZF1"

Query Match 83.9%; Score 473.4; DP 24; Length 2910;
Best Local Similarity 98.8%; Pred. No. 5.6e-116;
Matches 489; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY 1 GAGCTCTCTC-CCCCCAATCCCTCCGCTGAGGAACTACAAAGAGAAAAAAT 60
Db 2416 GAGCTCTCTCTCCCCCAATCCCTCCGCTGAGGAACTACAAAGAGAAAAAAT 2475
QY 61 TCCGGAAGACAGGATTAGAGGAAGCAAAATTCACCTGGTGCCCTTGGCTGCCGGGAAGT 120
Db 2476 TCCGGAAGACAGGATTAGAGGAAGCAAAATTCACCTGGTGCCCTTGGCTGCCGGGAAGT 2535
QY 121 GGACTAGAGAGTCTCGCGGCGAGCCCGGAGCCCGGCTTCCCGCGGCTTTAGCCCGGC 180
Db 2536 GGACTAGAGAGTCTCGCGGCGAGCCCGGAGCCCGGAGCCCGGCTTCCCGCGGCTTTAGCCCGGC 2595
QY 181 GGGCCCGGGGGGAAGGGGACGACGCGGAGCCCTAAGACACCTGCTGTACCCCTCC 240
Db 2596 GGGCCCGGGGGGAAGGGGACGACGCGGAGCCCTAAGACACCTGCTGTACCCCTCC 2655
QY 241 AC-----CCCCACCCACCCACCTCCCCCAACTCCCTAGATGTGTCTGTGGCGGCTGAA 295
Db 2656 ACCCCACCCACCCACCCACCCACCTCCCCCAACTCCCTAGATGTGTCTGTGGCGGCTGAA 2715
QY 296 CGTCGCCGTTTAAAGGGGGGGCGCCCGGCTCCACGTCCTTCTGCTGAGTACTGAACTA 355
Db 2716 CGTCGCCGTTTAAAGGGGGGGCGCCCGGCTCCACGTCCTTCTGCTGAGTACTGAACTA 2775
QY 356 CATAAACAGAGCGCGGAGACGGGGGGGAGGAGGAGACAGGCTTTGACCCGATAGT 415
Db 2776 CATAAACAGAGCGCGGAGACGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 2835
QY 416 AACCTCTGGCTCGGTGACGCCCAATCTATAAAGGAAGTACTCCCGGCAAAACCCCGT 475
Db 2836 AACCTCTGGCTCGGTGACGCCCAATCTATAAAGGAAGTACTCCCGGCAAAACCCCGT 2895
QY 476 AATTGCGAGCGAGAG 490
Db 2896 AATTGCGAGCGAGAG 2910

RESULT 7
AAC09615
ID AAC09615 standard; cDNA; 227 BP.
XX
XX AAC09615;
XX
XX 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 13690.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
XX Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 990S-0122487.
XX

```

```

XX PA (GSET) GENSET.
XX PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX PI WPI: 2003-500381/45.
XX DR
XX PI New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX XX
XX PS Claim 1: SEQ ID 13690; 71pp + CD-ROM; English.
XX XX
XX CC The present sequence is one of a large number of 5' ESTs derived from
XX CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX CC identified within the present sequence. The 5' ESTs were prepared from
XX CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX CC sequences usually correspond mainly to the 3' untranslated region (UTR)
XX CC of the mRNA because they are often obtained from oligo-dT primed cDNA
XX CC libraries. Such ESTs are not well suited for isolating cDNA sequences
XX CC derived from the 5' ends of mRNAs and even in those cases where longer
XX CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX CC They are used to obtain upstream regulatory sequences and to design
XX CC expression and secretion vectors.
XX XX
XX SQ Sequence 227 BP: 44 A; 65 C; 73 G; 45 T; 0 other;
Query Match 21.2%; Score 119.4; DB 21; Length 227;
Best Local Similarity 99.2%; Pred. No. 3.3e-22;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 444 ATAAAGGACCTAGTCCCGGCAAAACCCCGTAAATTCGAGCGAGTGAGTGGGCGCG 503
Db 1 ATAAAGGACCTAGTCCCGGCAAAACCCCGTAAATTCGAGCGAGTGAGTGGGCGCG 50

Oy 504 GACCCGCGAGCGAGCGAGCGCTCTCTCCGGGCTCGCGAGGCGGCGGAGCT 563
Db 61 GACCCGCGAGCGAGCGAGCGCTCTCTCCGGGCTCGCGAGGCGGCGGAGCT 120

Oy 564 C 564
Db 121 C 121

RESULT 8
AAD21326
XX ID AAD21326 standard; DNA; 7260 BP.
XX AC AAD21326;
XX DT
XX DT 28-JAN-2002 (first entry)
XX DE Human ATP binding cassette transporter 1 (ABCI) gene.
XX XX
KW Human; ATP binding cassette transporter 1; ABC1; coronary heart disease;
KW dermatologica; atherosclerosis; cardiovascular; inflammatory disease;
KW psoriasis; lipid disorder; antibacterial; septic shock; gene therapy;
KW immunosuppressive; lupus erythematosus; rheumatoid arthritis; ds.
XX XX
OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 321..7106
XX FT CDS /*tag= a
XX FT CDS /*tag= b
XX FT CDS /*tag= c
XX FT variation /note= "alternative open reading frame of AAI70314"
XX FT variation -replace(976,A)
XX FT variation replace(1516,C)
XX FT variation /*tag= d
XX FT variation replace(2969,G)

20-MAR-2000: 2000EP-0105820.
20-MAR-2000: 2000EP-0105820.
(FARB ) BAYER AG.
Schmitz G, Bodzioch M;
WPI: 2001-640388/74.
P-PSDB: AAE13022.
XX New adenosine triphosphate binding cassette transporter-1 gene
XX PT polymorphisms, useful for diagnosing and treating lipid disorders,
XX PT cardiovascular diseases and inflammatory diseases -
XX XX
XX PS Example 1: Fig 1: 48pp; English.
XX XX
XX CC The invention relates to four common polymorphisms in the gene encoding
XX CC ATP-binding cassette transporter-1 (ABCI). ABC1 is associated with
XX CC decreased ApCA-1 mediated efflux of cholesterol. The polymorphisms in
XX CC ABC1 directly affects cellular lipid homeostasis, which is a key factor
XX CC in the atherogenetic processes. The ABC1 polymorphisms are useful for
XX CC diagnosing and treating lipid disorders, cardiovascular diseases
XX CC (coronary heart disease, atherosclerosis) and inflammatory diseases
XX CC (psoriasis, lupus erythematosus). The identification of ABC1 as a
XX CC transporter for interleukin-beta (IL-beta) identifies this gene as
XX CC a candidate for treatment of inflammatory diseases including rheumatoid
XX CC arthritis and septic shock. The present sequence is human ABC1 gene.
XX XX
XX SQ Sequence 7260 BP: 1834 A; 1765 C; 1305 G; 1756 T; 0 other;
Query Match 17.4%; Score 98; DB 22; Length 7260;
Best Local Similarity 100.0%; Pred. No. 4.4e-16;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 467 AAACCCCTAATGGGAGCGAGAGTGAGTGGGCGGCGGAGCCGAGAGCGGAGCCG 526
Db 1 AAACCCCTAATGGGAGCGAGAGTGAGTGGGCGGCGGAGCCGAGAGCGGAGCCG 50

Oy 527 TTCCTCCCGGCTCGCGAGGCGGCGGCGGAGCTC 564
Db 61 TTCCTCCCGGCTCGCGAGGCGGCGGCGGAGCTC 98

RESULT 9
AAI70315
XX ID AAI70315 standard; cDNA; 7260 BP.
XX AC AAI70315;
XX DT
XX DT 07-JAN-2002 (first entry)
XX XX
XX DE Human ATP binding cassette transporter 1 (ABCI) cDNA.
XX XX
KW ATP binding cassette transporter 1; ABC1; human; lipid disorder;
KW cholesterol; cardiovascular disease; inflammatory disease;
KW antiinflammatory; antilipemic; antipsoriatic; dermatological;
KW Tangier disease; coronary heart disease; diagnosis; gene therapy;
KW polymorphism; ss.
XX XX
OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
XX FT CDS 321..7106
XX FT CDS /*tag= a
XX FT CDS 501..7106 /*tag= b
XX FT CDS /*tag= c
XX FT variation /note= "alternative open reading frame of AAI70314"
XX FT variation -replace(976,A)
XX FT variation /*tag= c
XX FT variation replace(1516,C)
XX FT variation /*tag= d
XX FT variation replace(2969,G)

```

variation /*tag= e
replace(3836,C)
/*tag= f

XX EP1136554-A1.

XX 26-SEP-2001.

XX 24-MAR-2000; 2000EP-C106401.

XX 24-MAR-2000; 2000EP-0106401.

XX (FARB) BAYER AG.

XX Schmitz G, Bodzioch M;

XX WPI; 2001-640389/74.

XX P-ESDB; AAM50228.

XX New adenosine triphosphate binding cassette transporter gene
polymorphisms, useful for diagnosing and treating lipid disorders,
PT cardiovascular diseases and inflammatory diseases -

XX Disclosure: Page 26-28; 41pp; English.

XX The present sequence is that of cDNA encoding the human adenosine
triphosphate (ATP) binding cassette transporter 1 (ABCI1) protein
CC (see AAM50227). The sequence includes an extended open reading
CC frame (ORF) to that provided by the sequence in AAI70314, using
CC an alternative ATG codon as initiation codon and thereby adding an
CC extra 40 N-terminal amino acids to the encoded ABCI1 protein (see
CC AAM50228). The invention provides 4 common polymorphisms in the
CC ABCI1 gene. These were identified by sequencing the ABCI1 gene in
CC different Tangier kindreds. In the variant genes (numbering as in
CC AAI70314), G is changed to A at position 586, T is changed to C at
CC position 1136, A is changed to G at position 2589 or G is changed
CC to C at position 3456, or any combination of these. All of these
CC polymorphisms alter the amino acid sequence of ABCI1 and therefore
CC may affect its function. The 2 most common polymorphisms (G596A)
CC and A2589G) are both associated with a decreased in vitro ApoA-I
CC mediated efflux of cholesterol from mononuclear phagocytes, a
CC feature typical of Tangier disease. 3 Of the variants (G596A,
CC A2589G and G3456C) are significantly increased in a population of
CC men having low high density lipoprotein-cholesterol levels and
CC established coronary heart disease (CHD) relative to CHD-free
CC control subjects. The use of the provided ABCI1 polymorphisms for
CC the diagnosis and treatment of lipid disorders, cardiovascular
CC diseases, and inflammatory diseases (e.g. psoriasis, lupus
CC erythematoses) is claimed. Modulation of ABCI1 transcripts or
CC proteins by antisense or ribozyme technology or RNA decoys is also
CC claimed.

XX Sequence 7260 BP; 1834 A; 1765 C; 1905 G; 1756 T; 0 other;

XX Query Match: 17.4%; Score 98; DB 22; Length 7260;

XX Best Local Similarity 100.0%; Pred. No. 4.4e-16;

XX Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 457 AAACCCCGGTAATTGGCAGGAGAGTGAGTGGCGGGGACCGGAGCGGAGCCGCCACCC 526

DB 1 AAACCCCGGTAATTGGCAGGAGAGTGAGTGGCGGGGACCGGAGCGGAGCGGAGCCGCCACCC 50

QY 527 TTCTCTCCCGGGCTGGCGGAGGCGAGGGCGGGGAGCTC 564

DB 61 TTCTCTCCCGGGCTGGCGGAGGCGAGGGCGGGGAGCTC 98

RESULT 10

XX AAD37268

XX ID AAD37268 standard; DNA; 221 BP.

XX AAD37268;

XX

21-AUG-2002 (first entry)

XX Human ABCI1 gene exon 1A.

XX Human; ATP-binding cassette 1; ABCI1 gene regulation; atherosclerosis;
KW cholesterol metabolism; hypercholesterolaemia; antisense therapy;
KW exon 1A; ds.

XX Homo sapiens.

XX WO200183746-A2.

XX 08-NOV-2001.

XX 02-MAY-2001; 2001WO-EP05488.

XX 02-MAY-2000; 2000US-2012807.

XX (AVEI) AVENTIS PHARMA SA.

XX Rosier-Montus M, Prades C, Lemoine C, Naudin L, Deneffe P;

XX Brewer B, Duverger N, Remaley A, Santamarina-Fojo S;

XX WP-; 2002-154404/20.

XX Isolated nucleic acid useful for modifying the ATP-binding cassette 1
PT (ABCI1) and screening for candidate modulatory compounds or substances

XX Claim 4; Page 132; 152pp; English.

XX The invention relates a nucleic acid which is capable of regulating the
transcription of human ATP-binding cassette 1 (ABCI1) gene, which is a
CC causal gene for pathologies linked to a dysfunctioning of cholesterol
CC metabolism, including diseases such as atherosclerosis. Polynucleotides
CC of the invention are used to screen candidate molecules or substances
CC that are capable of modulating the transcription of the ABCI1 gene. They
CC are used in antisense therapy. Compositions comprising sequences of the
CC invention are used to treat hypercholesterolaemia and atherosclerosis.
XX The present sequence is human ABCI1 gene exon 1A.

XX Sequence 221 BP; 44 A; 62 C; 73 G; 42 T; 0 other;

XX Query Match: 16.1%; Score 91; DB 24; Length 221;

XX Best Local Similarity 100.0%; Pred. No. 1.1e-14;

XX Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 GTAATTGGCAGGAGAGTGAGTGGCGGGGACCGGAGCGGAGCCGACCCCTCTCTC 533

DB 1 GTAATTGGCAGGAGAGTGAGTGGCGGGGACCGGAGCGGAGCCGACCCCTCTCTC 60

QY 534 CCGGGCTGGCGGAGGCGAGGGCGGGGAGCTC 564

DB 61 CCGGGCTGGCGGAGGCGAGGGCGGGGAGCTC 91

RESULT 11

XX AAH07432

XX ID AAH07432 standard; cDNA; 736 BP.

XX AC AAH07432;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:4267.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

```

XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPT: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 1; SEQ ID 4257; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 736 BP; 163 A; 199 C; 199 G; 170 T; 5 other;
XX
XX Query Match 15.9%; Score 89.4; DB 22; Length 736;
XX Best Local Similarity 98.9%; Pred. No. 4.3e-14;
XX Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 474 GTAATTGCGAGCAGAGTGTAGTGGCGCGGACCCGACGAGCCGACCCCTTCTCTC 533
XX ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1 GTAGTTGCGAGCAGAGTGTAGTGGCGCGGACCCGACGAGCCGACCCCTTCTCTC 60
XX
XX 534 CCGGGCTGCGGACGAGCAGCGGGGGAGCTC 564
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 61 CCGGGCTGCGGACGAGCAGCGGGGGAGCTC 91
XX
XX RESULT 12
XX AAH18606
XX ID AAH18606 standard: cDNA; 1556 BP.
XX AC AAH18606;
XX XX
XX 26-JUN-2001 (first entry)
XX DE Human cDNA sequence SEQ ID NO:18808.
XX XX
XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX XX

```

```

OS Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPT: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 18808; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 1556 BP; 380 A; 363 C; 399 G; 414 T; 0 other;
XX
XX Query Match 15.9%; Score 89.4; DB 22; Length 1556;
XX Best Local Similarity 98.9%; Pred. No. 5.4e-14;
XX Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 474 GTAATTGCGAGCAGAGTGTAGTGGCGCGGACCCGACGAGCCGACCCCTTCTCTC 533
XX ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 1 GTAGTTGCGAGCAGAGTGTAGTGGCGCGGACCCGACGAGCCGACCCCTTCTCTC 60
XX
XX 534 CCGGGCTGCGGACGAGCAGCGGGGGAGCTC 564
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 61 CCGGGCTGCGGACGAGCAGCGGGGGAGCTC 91
XX
XX RESULT 13
XX AAH04729
XX ID AAH04729 standard: cDNA; 763 BP.
XX AC AAH04729;
XX XX
XX 26-JUN-2001 (first entry)
XX

```

```
XX Human cDNA clone (5'-primer) SEQ ID NO:1564.
DE
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1C74617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INSI.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki I;
XX
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 1: SEQ ID 1564; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 763 32; 137 A; 205 C; 260 G; 158 T; 3 other;
XX
XX Query Match 15.8%; Score 89; DB 22; Length 763;
XX Best Local Similarity 100.0%; Pred. No. 5.6e-14;
XX Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 476 AATTCCGAGCGAGAGTGTAGTGGGCGCGGACCCGAGAGCGCGACCCCTTCTCTCC 535
Dd 1 AATTCCGAGCGAGAGTGTAGTGGGCGCGGACCCGAGAGCGCGACCCCTTCTCTCC 60
QY 536 GGGCTGCGGCGAGCGAGCGGGGGGAGCTC 564
Dd 61 GGGTGGCGGCGAGCGAGCGGGGGGAGCTC 89
RESULT 14
AAH17451
```

```
ID AAH17451 standard; cDNA; 1750 BP.
XX
AC AAH17451;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:16905.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EP1C74617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INSI.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki I;
XX
XX WPI: 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8: SEQ ID 16905; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 1750 BP; 291 A; 489 C; 586 G; 384 T; 0 other;
XX
XX Query Match 15.8%; Score 89; DB 22; Length 1750;
XX Best Local Similarity 100.0%; Pred. No. 7.2e-14;
XX Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 476 AATTCCGAGCGAGAGTGTAGTGGGCGCGGACCCGAGAGCGCGACCCCTTCTCTCC 535
Dd 1 AATTCCGAGCGAGAGTGTAGTGGGCGCGGACCCGAGAGCGCGACCCCTTCTCTCC 60
QY 536 GGGCTGCGGCGAGCGAGCGGGGGGAGCTC 564
AAH17451
```

Db 61 GGGCTCGGCGAGGCGAGGCGGGGAGGCTC 89

RESULT 15

AAK51683
ID AAK51683 standard; cDNA; 7281 BP.

AC AAK51683;

DT 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 228.

XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

XX PR 01-SEP-2000; 2000US-0654936.

XX PR 15-SEP-2000; 2000US-0663561.

XX PR 20-OCT-2000; 2000US-0693325.

XX PR 30-NOV-2000; 2000US-0728422.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

XX PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX PI Xue AZ, Yang Y, Wejhrman T, Goodrich R;

XX DR WPI; 2001-476283/51.

XX DR P-PSDB; AAM78550.

XX PT Nucleic acids encoding polypeptides with cytokine-like activities,

XX PT useful in diagnosis and gene therapy

XX PS Claim 1; Page 1086-1096; 6221pp; English.

XX CC The invention relates to polynucleotides (AAK51456-AAK51435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

XX CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666

XX CC (AAM80020) are omitted as the relevant pages from the sequence listing

XX CC were missing at the time of publication.

XX SQ Sequence 7281 BP; 1831 A; 1773 C; 1915 G; 1762 T; 0 other;

Query Match 13.7%; Score 77.2; DB 22; Length 7281;

Best Local Similarity 86.7%; Pred. No. 1.5e-10;

Matches 85; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 527 TTCTCTCCCGGCTGCGGCGAGGCGAGGCGGGGAGCTC 564
|||||
DB 82 TTCTCTCCCGGCTGCGGCGAGGCGAGGCGGGGAGCTC 119
|||||

Search completed: February 3, 2003, 16:30:12
Job time : 248.179 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 12:46:48 : Search time 96.1359 Seconds
(without alignments)
8245.650 Million cell updates/sec

Title: US-09-596-141c-3_COPY_1292_1643

Perfect score: 352

Sequence: 1 ggaacctaagacactctgtg.....cagggcaggcgaggagctc 352

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	352	100.0	1197	24	ABL58400 Human large ATP-bi
2	352	100.0	1643	22	AAF24581 Nucleotide sequenc
3	352	100.0	1643	22	AAF24703 Nucleotide sequenc
4	336	95.5	18399	22	AAF92831 Human ABC1 genomic
5	335.4	95.3	3231	24	AAD37265 Human ABC1 transcr
6	262.4	74.3	2910	24	AAD37267 Human ABC1 transcr
7	119.4	33.9	227	21	AAC09615 Human secreted pro
8	98	27.8	7260	22	AAD21326 Human ATP binding
9	98	27.8	7260	22	AAI70315 Human ATP binding

10	91	25.9	221	24	AAI37268	Human: ABC1 gene ex	
11	89.4	25.4	736	22	AAH07432	Human: CDNA clone (
12	89.4	25.4	1556	22	AAH18606	Human: cDNA sequenc	
13	89	25.3	763	22	AAH04729	Human: CDNA clone (
14	89	25.3	1750	22	AAH17451	Human: CDNA sequenc	
15	77.2	21.9	7281	22	AAK51683	Human polynucleoti	
16	75.4	21.4	7086	22	ABA09200	Human ABCA1 homolo	
17	75.4	21.4	7086	22	AAK52657	Human polynucleoti	
18	75	21.3	9854	22	AAI06121	Human ABC1 DNA seq	
19	67	19.0	10442	22	AAI24680	Nucleotide sequenc	
20	67	19.0	10442	22	AAI24702	Nucleotide sequenc	
21	60	17.0	10474	22	AAI24685	Nucleotide sequenc	
22	60	17.0	10474	22	AAI24686	Nucleotide sequenc	
23	60	17.0	10474	22	AAI24707	Nucleotide sequenc	
24	60	17.0	10474	22	AAI24708	Nucleotide sequenc	
25	32.2	10.9	114955	20	AAI53491	Human adenosine Al	
C	26	37.4	10.6	1599	16	AAQ97701	Rat melanocortin r
27	36.4	10.3	38	22	AAI93064	ABC1 polymorphism	
28	36.2	10.3	700	22	AAH92242	Human inflammatory	
C	29	35	9.9	486	23	AAI81261	DNA encoding novel
C	30	34.8	9.9	10732	21	AAAI0594	Gene encoding a su
C	31	34.4	9.8	669	23	AAI83553	DNA encoding novel
32	34.2	9.7	1610	22	AAI80546	Receptor #34 parti	
33	34	9.7	524	22	AAI92108	Human CDNA 5'-end	
34	34	9.7	524	22	AAK93547	Human CDNA clone r	
35	34	9.7	1769	22	AAK94827	Human full-length	
36	33.8	9.6	3168	24	ABQ91991	Human NF-kB activa	
37	33.8	9.6	3168	24	ABQ91992	Human NF-kB activa	
C	38	33.2	9.4	695	22	AAH04659	Human CDNA clone (
C	39	33.2	9.4	1235	22	AAH08064	Human extracellular
C	40	33.2	9.4	1877	22	AAH77797	Nucleotide sequenc
C	41	33.2	9.4	1877	22	AAI93858	Human stomach can
C	42	33.2	9.4	1877	22	AAH15566	Human CDNA sequenc
C	43	33	9.4	510	22	AAK88466	Human digestive sy
C	44	33	9.4	785	24	ABQ41592	Oligonucleotide fo
45	33	9.4	785	24	ABQ41593	Oligonucleotide fo	

ALIGNMENTS

RESULT :
ABL58400
ID ABL58400 standard; DNA; 1197 BP.
XX
AC ABL58400;
XX

30-JUL-2002 (first entry)

Human large ATP-binding cassette transporter 1 (hABC1) promoter sequence.
XX
DE Human large ATP-binding cassette transporter 1; ABC1; promoter;
XX
KW Human; large ATP-binding cassette transporter 1; ABC1; promoter;
XX
KW antiarteriosclerotic; gene transfer; transactivator; ds.
XX

OS Homo sapiens.

XX
PN WO200183506-A1.

XX
PD 08-NOV-2001.

XX
PF 27-APR-2001; 2001WO-US13654.

XX
PR 25-APR-2000; 2000US-0560372.

XX
PA (UYCO) UNIV COLUMBIA NEW YORK.

XX
PI Tall AR;

XX
DR WPI; 2002-049334/06.

XX
PT Novel isolated human large ATP-binding cassette transporter 1 promoter
XX
PT capable of directing transcription of heterologous coding sequence
XX
PT positioned downstream to it, useful for expressing foreign DNA in host

XX DE Nucleotide sequence of the 5' flanking region of the human ABC1 gene.
XX KW Human; adenosine triphosphate binding cassette protein 1; ABC1;
KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
XX atherosclerosis; cholesterol transport; ss.
XX OS Homo sapiens.
XX PN WC230078971-A2.
XX PC 28-DEC-2000.
XX PF 16-JUN-2000; 2000WO-US16591.
XX PR 18-JUN-1999; 93US-0140264.
XX PR 14-SEP-1995; 93US-0153872.
XX PR 19-NOV-1999; 99US-0166573.
XX PA (CVTE-) CV THERAPEUTICS INC.
XX PA (UNIW) UNIV WASHINGTON.
XX PI Lawn RM, Wade D, Oram JF, Garvin M;
XX WPI; 2001-1378:1/14.
XX AD Adenosine triphosphate (ATP) binding cassette protein (ABC) 1
PT polynucleotides and polypeptides, useful for treatment of heart disease
PT and other disorders associated with hypercholesterolemia and
PT atherosclerosis -
XX Disclosure; Page 138-139; 211pp; English.
XX CC The present sequence represents the 5' flanking region of the human
CC adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABC1
CC resides in cell membranes and utilises ATP hydrolysis to transport a wide
CC variety of substrates across the plasma membrane. ABC1 is a pivotal
CC protein in the apolipoprotein-mediated mobilisation of intracellular
CC cholesterol stores. ABC1 is defective in Tangier disease, a genetic
CC disorder characterised by abnormal HDL-cholesterol metabolism. The ABC1
CC gene is localised to chromosome 9q22-9q31. The ABC1 genes and proteins
CC are useful for developing pharmaceutical agents for the treatment of
CC heart disease and other disorders associated with hypercholesterolemia
CC and atherosclerosis. The genes are useful for developing screening assays
CC to screen for compounds that regulate the expression of genes associated
CC with cholesterol transport. The genes and proteins are also useful for
CC are also useful as diagnostic indicators of cardiovascular disease and
CC other disorders associated with hypercholesterolemia.
XX SQ Sequence 1643 BP; 370 A; 413 C; 457 G; 403 T; 0 other:
Query Match 100.0%; Score 352; DB 22; Length 1643;
Best Local Similarity 100.0%; Pred. No. 3.5e-89;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGACCTTAAGACACCTGCTGTACCTCCACCCACCCACCCACCTCCGCCCACTCCC 60
DB 1292 GGACCTTAAGACACCTGCTGTACCTCCACCCACCCACCCACCTCCGCCCACTCCC 1351
QY 61 TAGATGTGTCGGGGGCTGACGCTGCGCCGTTAAGGGGGGGGGGGGGGGGGGGGGGG 120
DB 1352 TAGATGTGTCGGGGGCTGACGCTGCGCCGTTAAGGGGGGGGGGGGGGGGGGGGG 1411
QY 121 CTTCTCTCTGAGTGACCTGACTACATAAAGAGGGGGGGGGGGGGGGGGGGGGGG 180
DB 1412 CTTCTCTCTGAGTGACCTGACTACATAAAGAGGGGGGGGGGGGGGGGGGGGG 1471
QY 181 GAGCACAGGCTTTGACCGATAGTAACCTCTGCGTCTGCTGAGCGGATCTATAAAGGA 240
DB 1472 GAGCACAGGCTTTGACCGATAGTAACCTCTGCGTCTGCTGAGCGGATCTATAAAGGA 1531
QY 241 ACTAGTCCCGGCATAAACCCTGTAAT7GGAGCGGAGAGTGAGTGGGGGGGGGGGGAGG 300

DB 1532 ACTAGTCCCGGCAAAACCCCGTAATTCGAGCGAGAGTGAGTGGGGCGGAGCCGCGAG 1591
QY 301 ACCGAGGCGGACCCCTTCTCTCCCGGGCTGCGGCGAGGCGAGGCGGGAGCTC 352
DB 1592 AGCGGAGCGGACCCCTCTCTCCCGGGCTGCGGCGAGGCGAGGCGGGGAGCTC 1643
RESULT 4
AAF92831 standard; DNA; 183999 BP.
XX AAF92831;
XX DT 17-MAY-2001 (first entry)
XX DE Human ABC1 genomic DNA.
XX KW High density lipoprotein-cholesterol; HDL-C; cardiovascular; ABC1; ds.
XX OS Homo sapiens.
XX PN WO200115676-A2.
XX PD 08-MAR-2001.
XX PF 01-SEP-2000; 2000WO-IB01492.
XX PR 01-SEP-1999; 99US-0151977.
XX PR 15-MAR-2000; 2000US-0526193.
XX PR 23-JUN-2000; 2000US-0213958.
XX PA (UYBR-) UNIV BRITISH COLOMBIA.
XX PA (XENO-) XENON GENETICS INC.
XX PI Hayden MR, Brooks-Wilson AR, Pimstone SN, Clee SM;
XX WPI; 2001-244356/25.
XX PT Treating a lower than normal high density lipoprotein-cholesterol
PT (HDL-C) level, a higher than normal triglyceride level, or a
PT cardiovascular disease, by administering a compound that modulates LXR-
PT or RXR-mediated transcriptional activity -
XX Claim 8; Fig 1; 317pp; English.
XX CC The present invention relates to a method for treating a patient
CC diagnosed as having a lower than normal high density
CC lipoprotein-cholesterol (HDL-C) level, a higher than normal
CC triglyceride level, or a cardiovascular disease, involving
CC administering a compound that modulates LXR- or RXR-mediated
CC transcriptional activity or ABC1 expression or activity.
CC The LXR gene product may be used in an assay to identify
CC compounds useful for the treatment of a disease or condition selected a
CC lower than normal HDL cholesterol level, a higher than normal
CC triglyceride level, and a cardiovascular disease.
XX SQ Sequence 183999 BP; 49549 A; 37944 C; 41170 G; 54950 T; 386 other:
Query Match 95.5%; Score 336; DB 22; Length 183999;
Best Local Similarity 98.3%; Pred. No. 3.4e-84;
Matches 351; Conservative 0; Mismatches 1; Indels 5; Gaps 1;
QY 1 GGACCTTAAGACACCTGCTGTACCTCCACCCACCCACCCACCTCCGCCCACTCCC 55
DB 28524 GGACCTTAAGACACCTGCTGTACCTCCANNNCNNCCACCCACCCACCTCCGCCCA 28583
QY 56 CTCCTAGATGTGTGTGGGGGGCTGAGCTCGCCCGTTTAAAGGGGGGGGGGGGGGGCTC 115
DB 28584 CTCCTAGATGTGTGTGGGGGGGGCTGAGCTCGCCCGTTTAAAGGGGGGGGGGGGGCTC 28643
QY 116 ACSTGCTTTCTGCTGAGTGACTGAACTACATAACAGAGCCCGGAGGAGGGGGGAGG 175

```
Db 28644 ACCTGCTTTCTGCTGAGTGACTGAACATACATAAACAGAGGCGCGGAACGCGGGGGGAGG 28703
QY 176 AGGAGACACAGCGCTTGGACCGANTGATACCTCTCGGCTCGGTGAGCGGCAATCTATAA 235
Db 28704 AGGAGACACAGCGCTTGGACCGATAGTAACCTCTCGGCTCGGTGAGCGGCAATCTATAA 28763
QY 236 AAGGAACCTAGTCCCGGCAAAAACCCCGTAATTGGAGCGAGAGTGAGTGGGGCGGGGACC 295
Db 28764 AAGGAACCTAGTCCCGGCAAAAACCCCGTAATTGGAGCGAGAGTGAGTGGGGCGGGGACC 29823
QY 296 CGGAGAGCGGAGCGCGACCTCTCTCTCCGGGCTGCGGCGAGGCGGCGGAGCTC 352
Db 28824 CGGAGAGCGGAGCGCGACCTCTCTCTCCGGGCTGCGGCGAGGCGGCGGAGCTC 28880

RESULT 5
AAD37265
ID AAD37265 standard; DNA: 3231 BP.
XX
AC AAD37265;
XX
D" 21-AUG-2002 (first entry)
XX
DE Human ABC1 transcription regulatory DNA #1.
XX
KW Human; ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
KW cholesterol metabolism; hypercholesterolaemia; antisense therapy; ds.
XX
OS Homo sapiens.
XX
PN W020183746-A2.
XX
PD 08-NOV-2001.
XX
PF 02-MAY-2001; 2001WO-EP05488.
XX
PR 02-MAY-2000; 2000US-201280P.
XX
PA (AVET ) AVENTIS PHARMA SA.
XX
PI Rosier-Montus M, Prades C, Lemoine C, Naudin L, Deneffe P;
PI Brewer B, Duverger N, Remaley A, Santamarina-Fojo S;
XX
DR WP; 2002-154404/20.
XX
PT Isolated nucleic acid useful for modifying the ATP-binding cassette 1
PT (ABC1) and screening for candidate modulatory compounds or substances
XX
PS Claim 1; Page 130-131; 152pp; English.
XX
CC The invention relates a nucleic acid which is capable of regulating the
CC transcription of human ATP-binding cassette 1 (ABC1) gene, which is a
CC causal gene for pathologies linked to a dysfunctioning of cholesterol
CC metabolism, including diseases such as atherosclerosis. Polynucleotides
CC of the invention are used to screen candidate molecules or substances
CC that are capable of modulating the transcription of the ABC1 gene. They
CC are used in antisense therapy. Compositions comprising sequences of the
CC invention are used to treat hypercholesterolaemia and atherosclerosis.
CC The present sequence is human ABC1 transcription regulating DNA.
XX
SQ Sequence 3231 BP; 809 A; 773 C; 876 G; 773 T; 0 other;
Q-ery Match 95.3%; Score 335.4; DB 24; Length 3231;
Best Local Similarity 98.3%; Pred. No. 1.9e-84;
Matches 351; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

Qy 1 GGACCTAAGACACTGCTGTACCTTCAC-----CCACCCACCCACCCACCTCCGCCAA 55
Db 2628 GGACCTAAGACACTGCTGTACCTTCACCCACCCACCCACCCACCCACCTCCGCCAA 2687
Qy 56 CTCCTAGATGTGCTGGGGGCTGACGTGCGCCGTTTAAGGGGCGGCGCGGCTCC 115
```

```
Db 2688 CTCCTAGATGTGCTGGGGGCGGCTGAACGTGCGCCGCTTTAAGGGGCGGCGCGGCTCC 2747
QY 116 ACCTGCTTTCTGCTGAGTGACTGAACATACATAAACAGAGGCGCGGCAATCTATAA 175
Db 2748 ACCTGCTTTCTGCTGAGTGACTGAACATACATAAACAGAGGCGCGGCAATCTATAA 2807
QY 176 AGGAGACACAGCGCTTGGACCGATAGTAACCTCTCGGCTCGGTGAGCGGCAATCTATAA 235
Db 2808 AGGAGACACAGCGCTTGGACCGATAGTAACCTCTCGGCTCGGTGAGCGGCAATCTATAA 2867
QY 236 AAGGAACCTAGTCCCGGCAAAAACCCCGTAATTGGAGCGAGAGTGAGTGGGGCGGGGACC 295
Db 2868 AAGGAACCTAGTCCCGGCAAAAACCCCGTAATTGGAGCGAGAGTGAGTGGGGCGGGGACC 2927
QY 296 CGGAGAGCGGAGCGCGACCTCTCTCTCCGGGCTGCGGCGAGGCGGCGGAGCTC 352
Db 2928 CGGAGAGCGGAGCGCGACCTCTCTCTCCGGGCTGCGGCGAGGCGGCGGAGCTC 2984

RESULT 6
AAD37267
ID AAD37267 standard; DNA: 2910 BP.
XX
AC AAD37267;
XX
D" 21-AUG-2002 (first entry)
XX
DE Human ABC1 transcription regulatory DNA #3.
XX
KW Human; ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
KW cholesterol metabolism; hypercholesterolaemia; antisense therapy; ds.
XX
OS Homo sapiens.
XX
FE Key Location/Qualifiers
FT protein_bind 12..23
FT /*tag= a
FT /bound_moiety= "LMO2COM/MYOD"
FT protein_bind 97..107
FT /*tag= b
FT /bound_moiety= "DeltaE1"
FT protein_bind 110..125
FT /*tag= c
FT /bound_moiety= "S8/NKX2.5"
FT protein_bind 196..211
FT /*tag= d
FT /bound_moiety= "S8"
FT protein_bind 228..237
FT /*tag= e
FT /bound_moiety= "GATA"
FT protein_bind 399..410
FT /*tag= f
FT /bound_moiety= "IK2"
FT protein_bind 412..420
FT /*tag= g
FT /bound_moiety= "LYF1"
FT protein_bind 528..539
FT /*tag= h
FT /bound_moiety= "LMO2COM/MYOD/DeltaE1"
FT protein_bind 549..556
FT /*tag= i
FT /bound_moiety= "LYF1"
FT protein_bind 558..568
FT /*tag= j
FT /bound_moiety= "DeltaE1"
FT protein_bind 590..596
FT /*tag= k
FT /bound_moiety= "NKX2.5"
FT protein_bind 608..620
FT /*tag= l
FT /bound_moiety= "NRY/CAA"
FT protein_bind 708..715
FT /*tag= m
```

```
FT protein_bind /bound_moiety= "MZFl"
FT 723..730
FT /*tag= n
FT /bound_moiety= "MZFl"
FT 771..785
FT /*tag= o
FT /bound_moiety= "HfH2/SRY/evil"
FT 803..812
FT /*tag= p
FT /bound_moiety= "CREBFl/VBP"
FT 831..837
FT /*tag= r
FT /bound_moiety= "NKX2.5"
FT 1076..1089
FT /*tag= s
FT /bound_moiety= "GATA"
FT 1173..1188
FT /*tag= t
FT /bound_moiety= "LXRalpha/deltaEF1"
FT 1189..1209
FT /*tag= u
FT /bound_moiety= "DeltaEF1/LXFl/IK2"
FT 1483..1491
FT /*tag= v
FT /bound_moiety= "AP4"
FT 1498..1514
FT /*tag= w
FT /bound_moiety= "LMO2-COM/MYOD/deltaEF1/E47"
FT 1524..1545
FT /*tag= x
FT /bound_moiety= "ZID/deltaEF1"
FT 1597..1607
FT /*tag= y
FT /bound_moiety= "DeltaEF1"
FT 1622..1627
FT /*tag= z
FT /bound_moiety= "PPAR"
FT 1632..1637
FT /*tag= aa
FT /bound_moiety= "PPAR"
FT 1685..1698
FT /*tag= ab
FT /bound_moiety= "USF/NMYC/MYCMAX"
FT 1767..1797
FT /*tag= ac
FT /bound_moiety= "DeltaEF1"
FT 1809..1819
FT /*tag= ad
FT /bound_moiety= "DeltaEF1"
FT 1822..1833
FT /*tag= ae
FT /bound_moiety= "SRY"
FT 1840..1850
FT /*tag= af
FT /bound_moiety= "AP1"
FT 1942..1956
FT /*tag= ag
FT /bound_moiety= "HNF3beta"
FT 1978..1985
FT /*tag= ah
FT /bound_moiety= "NKX2.5"
FT 2008..2016
FT /*tag= ai
FT /bound_moiety= "PPAR/NKX2.5/PPAR"
FT 2019..2024
FT /*tag= aj
FT /bound_moiety= "PPAR/NKX2.5/PPAR"
FT 2051..2059
FT /*tag= ak
FT /bound_moiety= "GATA"
FT 2104..2111
FT /*tag= al
FT /bound_moiety= "SOX5"
FT 2114..2152
FT /*tag= am
FT /bound_moiety= "SRY/HFH/HNF3beta"
FT 2221..2228
FT /*tag= an
FT /bound_moiety= "MZFl"
FT 2234..2249
FT /*tag= ao
FT /bound_moiety= "IK2/NFkappaB/CREL"
FT 2259..2272
FT /*tag= ap
FT /bound_moiety= "LMO2COM/GATA"
FT 2289..2306
FT /*tag= aq
FT /bound_moiety= "MZFl/SRY"
FT 2313..2318
FT /*tag= ar
FT /bound_moiety= "PPAR"
FT 2321..2326
FT /*tag= as
FT /bound_moiety= "PPAR"
FT 2335..2342
FT /*tag= at
FT /bound_moiety= "MZFl"
FT 2361..2384
FT /*tag= au
FT /bound_moiety= "HNF3beta/SRY/evil"
FT 2426..2433
FT /*tag= av
FT /bound_moiety= "MZFl"
FT 2442..2451
FT /*tag= aw
FT /bound_moiety= "AP4"
FT 2455..2466
FT /*tag= ax
FT /bound_moiety= "SRY"
FT 2491..2498
FT /*tag= ay
FT /bound_moiety= "STAT"
FT 2524..2534
FT /*tag= az
FT /bound_moiety= "STAT/PPAR"
FT 2536..2541
FT /*tag= ba
FT /bound_moiety= "PPAR"
FT 2589..2600
FT /*tag= bb
FT /bound_moiety= "AP2"
FT 2610..2617
FT /*tag= bc
FT /bound_moiety= "MZFl"
FT 2634..2648
FT /*tag= bd
FT /bound_moiety= "LMO2COM/MYOD/E47"
FT 2657..2672
FT /*tag= be
FT /bound_moiety= "RREB1"
FT 2680..2698
FT /*tag= bf
FT /bound_moiety= "MZFl/CMVB"
FT 2728..2740
FT /*tag= bg
FT /bound_moiety= "SP1/GC"
FT 2743..2757
FT /*tag= bh
FT /bound_moiety= "JSF/NMYC/ARNT"
FT 2758..2773
FT /*tag= bi
FT /bound_moiety= "NFE2AP1"
FT 2774..2787
FT /*tag= bj
FT /bound_moiety= "XFD1/HFH"
FT 2794..2806
```

```

FT      /tag= bk
FT      /bound_moiety= "GC/SP1/MZF1"

Query Match      74.3%; Score 261.4; DB 24; Length 2910;
Best Local Similarity 97.9%; Pred. No. 1.1e-63;
Matches 277; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY      1  GGACCTAAGACACCTGCTGACCTCCAC-----CCACACCCACCCACCTCCCCCAA 55
DB      2628 GGACCTAAGACACCTGCTGACCTCCACCCACCCACCCACCCACCTCCCCCAA 2687

QY      55  CTCCTAGATGTCCTGGGGGGCTCAAGCTCCCGCTTAAAGGGGGGGGGCCCGGCTCC 115
DB      2688 CTCCTAGATGTCCTGGGGGGCTCAAGCTCCCGCTTAAAGGGGGGGGGCCCGGCTCC 2747

QY      116 ACCTGCTTCTGCTGAGTGAATACATAAAGAGAGCGCGGGAACGGGGGGGGAGG 175
DB      2748 ACCTGCTTCTGCTGAGTGAATACATAAAGAGAGCGCGGGAACGGGGGGGGAGG 2807

QY      176 AGGAGAGACAGCGCTTIGACCGATAGTACCTCTCGGCTCGGTGCGAGCGGAATCTATA 235
DB      2808 AGGAGAGACAGCGCTTIGACCGATAGTACCTCTCGGCTCGGTGCGAGCGGAATCTATA 2867

QY      236 AGGAACACTAGTCCCGGCAAAACCCCGTAATTCGGAGCGAGAG 278
DB      2868 AGGAACACTAGTCCCGGCAAAACCCCGTAATTCGGAGCGAGAG 2910

RESULT 7
AAC09615
ID      AAC09615 standard; cDNA; 227 BP.
XX
AC      AAC09615;
XX
DT      06-OCT-2000 (first entry)
XX
DE      Human secreted protein 5' EST, SEQ ID NO: 13690.
KW      Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW      gene therapy; chromosome mapping; ss.
XX
OS      Homo sapiens.
XX
PN      EP1033401-A2.
XX
PD      06-SEP-2000.
XX
PF      21-FEB-2000; 2000EP-0200610.
XX
PR      26-FEB-1999; 99US-0122487.
XX
PA      (GIST ) GENSET.
XX
PI      Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR      WPI: 2000-500381/45.
XX
PT      New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT      obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT      diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS      Claim 1; SEQ ID 13690; 71pp + CD-ROM; English.
XX
CC      The present sequence is one of a large number of 5' ESTs derived from
CC      mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC      identified within the present sequence. The 5' ESTs were prepared from
CC      total human RNAs or poly(A) RNAs derived from 30 different tissues. EST
CC      sequences usually correspond mainly to the 3' untranslated region (UTR)
CC      of the mRNA because they are often obtained from oligo-dT primed cDNA
CC      libraries. Such ESTs are not well suited for isolating cDNA sequences
CC      derived from the 5' ends of mRNAs and even in those cases where longer
CC      cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC      5' ESTs are derived from mRNAs with intact 5' ends and can therefore be

```

```

CC      used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC      in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC      They are used to obtain upstream regulatory sequences and to design
CC      expression and secretion vectors.
XX
SQ      Sequence 227 BP; 44 A; 65 C; 73 G; 45 T; 0 other;

Query Match      33.5%; Score 119.4; DB 21; Length 227;
Best Local Similarity 99.2%; Pred. No. 4.6e-24;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      232 ATAAAGGAAGTAAGTCCCGGCAAAACCCCGTAATTCGAGCGAGAGTGAAGTGGGCGCG 291
DB      1  ATAAAGGAAGTAAGTCCCGGCAAAACCCCGTAATTCGAGCGAGAGTGAAGTGGGCGCG 60

QY      292 GACCCGAGAGCGGAGCCGACCCCTTCTCTCCCGGGCTGCGGCGAGGCGAGCGGGAGCT 351
DB      61  GACCCGAGAGCGGAGCCGACCCCTTCTCTCCCGGGCTGCGGCGAGGCGAGCGGGAGCT 420

QY      352 C 352
DB      121 C 121

RESULT 8
AAD21326
ID      AAD21326 standard; DNA; 7260 BP.
XX
AC      AAD21326;
XX
DT      28-JAN-2002 (first entry)
XX
DE      Human ATP binding cassette transporter 1 (ABCL1) gene.
KW      Human; ATP binding cassette transporter 1; ABCL1; coronary heart disease;
KW      dermatological; atherosclerosis; cardiovascular; inflammatory disease;
KW      psoriasis; lipid disorder; antibacterial; septic shock; gene therapy;
KW      immunosuppressive; lupus erythematosus; rheumatoid arthritis; ds.
XX
OS      Homo sapiens.
XX
PN      Key
XX      Location/Qualifiers
XX      321..7106
XX      CDS
XX      /*tag= a
XX      /product= "Human ABC1 protein"
XX
PF      EP1136552-A1.
XX
PD      26-SEP-2001.
XX
PR      20-MAR-2000; 2000EP-0105820.
XX
PA      (FARB ) BAYER AG.
XX
PI      Schmitz G, Bodzioch M;
XX
DR      WPI: 2001-640388/74.
XX
PS      P-PSDB; AAEL3022.
XX
CC      New adenosine triphosphate binding cassette transporter-1 gene
CC      polymorphisms, useful for diagnosing and treating lipid disorders,
CC      cardiovascular diseases and inflammatory diseases -
XX
PS      Example 1; Fig 1; 48pp; English.
XX
CC      The invention relates to four common polymorphisms in the gene encoding
CC      ATP-binding cassette transporter-1 (ABCL1). ABCL1 is associated with
CC      decreased ApoA1 mediated efflux of cholesterol. The polymorphisms in
CC      ABCL1 directly affects cellular lipid homeostasis, which is a key factor
CC      in the atherogenic processes. The ABCL1 polymorphisms are useful for
CC      diagnosing and treating lipid disorders, cardiovascular diseases

```

CC (coronary heart disease, atherosclerosis) and inflammatory diseases
 CC (psoriasis, lupus erythematosus). The identification of ABC1 as a
 CC transporter for interleukin-beta (IL-1beta) identifies this gene as
 CC a candidate for treatment of inflammatory diseases including rheumatoid
 CC arthritis and septic shock. The present sequence is human ABC1 gene.
 XX
 SQ Sequence 7260 BP; 1834 A; 1765 C; 1905 G; 1756 T; 0 other;
 Query Match: 27.8%; Score 98; DB 22; Length 7260;
 Best Local Similarity 100.0%; Pred. No. 1.1e-17;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 255 AAACCCCGTAATTCGAGCGAGAGTGAAGTGGCGCGGACCCGCGAGCGCGAGCCGACCC 314
 DB 1 AAACCCCGTAATTCGAGCGAGAGTGAAGTGGCGCGGACCCGCGAGCGCGAGCCGACCC 60
 QY 315 TTCTCTCCCGGGCTCGGCGAGCGCGAGCGCGGAGGCTC 352
 DB 61 TTCTCTCCCGGGCTCGGCGAGCGCGAGCGCGGAGGCTC 98
 RESULT 9
 ID AAI70315
 AC AAI703:5;
 XX
 DT 07-JAN-2002 (first entry)
 DE Human ATP binding cassette transporter 1 (ABC1) cDNA.
 XX
 KW ATP binding cassette transporter 1; ABC1; human; lipid disorder;
 KW cholesterol; cardiovascular disease; inflammatory disease;
 KW antiinflammatory; arthropathic; antipsoriatic; dermatological;
 KW Tangier disease; coronary heart disease; diagnosis; gene therapy;
 KW polymorphism; ss.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 CDS 321..7106
 FT /*tag= a
 CDS 501..7106
 FT /*tag= b
 FT /*note= "alternative open reading frame of AAI70314"
 FT replace(976,A)
 FT /*tag= c
 FT variation replace(1516,C)
 FT /*tag= d
 FT variation replace(2969,G)
 FT /*tag= e
 FT variation replace(3836,C)
 FT /*tag= f
 XX
 PF EPI136554-AL.
 PN
 PD 26-SEP-2001.
 XX
 PF 24-MAR-2000; 2000EP-0106401.
 XX
 PR 24-MAR-2000; 2000EP-0106401.
 XX
 PA (FARB) BAYER AG.
 XX
 XX Schmitz G, Bodzioch M;
 XX
 XX WPI: 2001-640369/74.
 DR P-PSDB; AAM50228.
 D3
 XX New adenosine triphosphate binding cassette transporter gene
 PT polymorphisms, useful for diagnosing and treating lipid disorders,
 PT cardiovascular diseases and inflammatory diseases
 XX

PS Disclosure; Page 26-28; 4:pp; English.
 XX
 CC The present sequence is that of cDNA encoding the human adenosine
 CC triphosphate (ATP) binding cassette transporter 1 (ABC1) protein
 CC (see AAM50227). The sequence includes an extended open reading
 CC frame (ORF) to that provided by the sequence in AAI70314, using
 CC an alternative ATG codon as initiation codon and thereby adding an
 CC extra 40 N-terminal amino acids to the encoded ABC1 protein (see
 CC AAM50228). The invention provides 4 common polymorphisms in the
 CC ABC1 gene. These were identified by sequencing the ABC1 gene in
 CC different Tangier kindreds. In the variant genes (numbered as in
 CC AAI70314), G is changed to A at position 596, T is changed to C at
 CC position 1136, A is changed to G at position 2589 or G is changed
 CC to C at position 3456, or any combination of these. All of these
 CC polymorphisms alter the amino acid sequence of ABC1 and therefore
 CC may affect its function. The 2 most common polymorphisms (G596A)
 CC and A2589G) are both associated with a decreased in vitro ApoA-1
 CC mediated efflux of cholesterol from mononuclear phagocytes, a
 CC feature typical of Tangier disease. 3 Of the variants (G596A,
 CC A2589G and G3456C) are significantly increased in a population of
 CC men having low high density lipoprotein-cholesterol levels and
 CC established coronary heart disease (CHD) relative to CHD-free
 CC control subjects. The use of the provided ABC1 polymorphisms for
 CC the diagnosis and treatment of lipid disorders, cardiovascular
 CC diseases, and inflammatory diseases (e.g. psoriasis, lupus
 CC erythematosus) is claimed. Modulation of ABC1 transcripts or
 CC proteins by antisense or ribozyme technology or RNA decoys is also
 CC claimed.
 XX
 SQ Sequence 7260 BP; 1834 A; 1765 C; 1905 G; 1756 T; 0 other;
 Query Match: 27.8%; Score 98; DB 22; Length 7260;
 Best Local Similarity 100.0%; Pred. No. 1.1e-17;
 Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 255 AAACCCCGTAATTCGAGCGAGAGTGAAGTGGCGCGGACCCGCGAGCGCGAGCCGACCC 314
 DB 1 AAACCCCGTAATTCGAGCGAGAGTGAAGTGGCGCGGACCCGCGAGCGCGAGCCGACCC 60
 QY 315 TTCTCTCCCGGGCTCGGCGAGCGCGAGCGCGGAGGCTC 352
 DB 61 TTCTCTCCCGGGCTCGGCGAGCGCGAGCGCGGAGGCTC 98
 RESULT 10
 ID AAI70315
 AC AAI703:5;
 XX
 DT 21-AUG-2002 (first entry)
 DE Human ABC1 gene exon 1A.
 XX
 KW Human; ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
 KW cholesterol metabolism; hypercholesterolaemia; antisense therapy;
 KW exon 1A; ds.
 XX
 CS Homo sapiens.
 PN WC200183746-A2.
 XX
 PD 08-NOV-2001.
 XX
 PF 02-MAY-2001; 2001WO-EP05488.
 XX
 PR 02-MAY-2000; 2000US-201280P.
 XX
 PA (AVET) AVENTIS PHARMA SA.
 XX
 XX Rosier-Montus M, Prades C, Lemoine C, Naudin L, Deneffe P;
 PI B-rewer B, Duverger N, Remaley A, Santamarina-Pojo S;
 XX

DR WPI; 2002-154404/20.
XX
XX Isolated nucleic acid useful for modifying the ATP-binding cassette 1
PT (ABC1) and screening for candidate modulatory compounds or substances
PT
PT
PT
PS Claim 4; Page 132; 152pp; English.
XX
XX The invention relates a nucleic acid which is capable of regulating the
CC transcription of human ATP-binding cassette 1 (ABC1) gene, which is a
CC causal gene for pathologies linked to a dysfunctioning of cholesterol
CC metabolism, including diseases such as atherosclerosis. Polynucleotides
CC of the invention are used to screen candidate molecules or substances
CC that are capable of modulating the transcription of the ABC1 gene. They
CC are used in antisense therapy. Compositions comprising sequences of the
CC invention are used to treat hypercholesterolemia and atherosclerosis.
XX The present sequence is human ABC1 gene exon 1A.
XX
SQ Sequence 221 BP; 44 A; 62 C; 73 G; 42 T; 0 other;
Query Match 25.9%; Score 91; DB 24; Length 221;
Best Local Similarity 100.0%; Pred. No. 4.3e-16;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 262 GTAATGCGAGCGAGAGTGGGGCGGGACCGCGAGCGAGCGAGCGACCTTCTCTC 321
DB 1 GTAATGCGAGCGAGAGTGGGGCGGGACCGCGAGCGAGCGAGCGACCTTCTCTC 60
QY 322 CCGGGTGGCGAGCGAGCGAGCGGGGAGCTC 352
DB 61 CCGGGTGGCGAGCGAGCGAGCGGGGAGCTC 91
RESULT 11
AAH07432
ID AAH07432 standard; cDNA; 736 BP.
XX
XX AAH07432;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA clone (5'-primer) SEQ ID NO:4267.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EF1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PI full-length cDNAs defined in the specification, and for the detection
PI and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 1; SEQ ID 4267; 2537pp + CD ROM; English.
XX

CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13532
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 736 BP; 163 A; 199 C; 199 G; 170 T; 5 other;
Query Match 25.4%; Score 89.4; DB 22; Length 736;
Best Local Similarity 98.9%; Pred. No. 1.6e-15;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 262 GTAATGCGAGCGAGAGTGGGGCGGGACCGCGAGCGAGCGAGCGACCTTCTCTC 321
DB 1 GTAATGCGAGCGAGAGTGGGGCGGGACCGCGAGCGAGCGAGCGACCTTCTCTC 60
QY 322 CCGGGTGGCGAGCGAGCGGGGAGCTC 352
DB 61 CCGGGTGGCGAGCGAGCGGGGAGCTC 91
RESULT 12
AAH18606
ID AAH18606 standard; cDNA; 1556 BP.
XX
XX AAH18606;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:18808.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EF1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PI full-length cDNAs defined in the specification, and for the detection
PI and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX Claim 1; SEQ ID 4267; 2537pp + CD ROM; English.
XX

PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -

XX Claim 8: SEQ ID 18808; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 1556 BP; 380 A; 363 C; 399 G; 414 T; 0 other;

Query Match 25.4%; Score 89.4; DB 22; Length 1556;
 Best Local Similarity 98.9%; Pred. No. 1.9e-15;
 Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 262 GTAATTGGGAGGAGTGTAGTGGGCGGGAGCCGACAGCCGAGCCGACCTTCTCTC 321

Db 1 GTAGTTGGGAGGAGTGTAGTGGGCGGGAGCCGACAGCCGAGCCGACCTTCTCTC 60

QY 322 CCGGGCTGGGAGGAGGCGGGGGGAGCTC 352

Db 61 CCGGGCTGGGAGGAGGCGGGGGGAGCTC 91

RESULT 13

AAH04729

ID AAH04729 standard; cDNA; 763 BP.

XX AC AAH04729;

XX 26-JUN-2001 (first entry)

DE Human cDNA clone (5'-primer) SEQ ID NO:1564.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000EP-0116776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -

PS Claim 1: SEQ ID 1564; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX Sequence 763 BP; 137 A; 205 C; 260 G; 158 T; 3 other;

Query Match 25.3%; Score 89; DB 22; Length 763;
 Best Local Similarity 100.0%; Pred. No. 2.1e-15;

Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 264 AATTGGGAGGAGTGTAGTGGGCGGGAGCCGACAGCCGAGCCGACCTTCTCTCC 323

Db 1 AATTGGGAGGAGTGTAGTGGGCGGGAGCCGACAGCCGAGCCGACCTTCTCTCC 60

QY 324 GGGCTGGGAGGAGGCGGGGGGAGCTC 352

Db 61 GGGCTGGGAGGAGGCGGGGGGAGCTC 89

RESULT 14

AAH17451

ID AAH17451 standard; cDNA; 1750 BP.

XX AC AAH17451;

XX 26-JUN-2001 (first entry)

DE Human cDNA sequence SEQ ID NO:16905.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0116776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.


```
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki I;
XX DR WPI; 2001-3-8749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX PS Claim 8; SEQ ID 16305: 2537bp + CD ROM; English.
XX CC The present invention describes primer sets for synthesizing 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences. AAH92446 to
XX CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.
XX SQ Sequence 1750 BP: 291 A; 489 C; 566 G; 384 T; 0 other;
Query Match 25.3%; Score 89; DB 22; Length 1750;
Best Local Similarity 100.0%; Pred. No. 2.6e-15;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 264 AATTGGCAGGAGAGTGGTGGGCGGGACCCGACAGCGAGCGAGCCCTCTCTCC 323
DB 1 AATTGGCAGGAGAGTGGTGGGCGGGACCCGACAGCGAGCGAGCCCTCTCTCC 50
QY 324 GGGCTGGCGCAGGCGAGGGCGGGGAGCTC 352
DB 61 GGGCTGGCGCAGGCGAGGGCGGGGAGCTC 89
RESULT 15
AAK51683
ID AAK51683 standard; cDNA; 7281 BP.
XX AC AAK51683;
XX XX
DT 06-NOV-2001 (first entry)
DE Human polynucleotide SEQ ID NO 228.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation; ss.
XX Homo sapiens.
XX WO200157150-A2.
XX 09-AUG-2001.
```

```
XX PF 05-FEB-2001; 2001WO-US04098.
XX PR 03-FEB-2000; 2000US-0496914.
XX PR 27-APR-2000; 2000US-0560875.
XX PR 20-JUN-2000; 2000US-0598075.
XX PR 19-JUL-2000; 2000US-0620325.
XX PR 01-SEP-2000; 2000US-0654936.
XX PR 15-SEP-2000; 2000US-0663561.
XX PR 20-OCT-2000; 2000US-0693325.
XX PR 30-NOV-2000; 2000US-0728422.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX PI Xue AZ, Yang Y, Wejhrman I, Goodrich R;
XX DR WPI; 2001-476283/51.
XX DR P-PSDB; AAK78550.
XX CC Nucleic acids encoding polypeptides with cytokine-like activities,
XX CC useful in diagnosis and gene therapy -
XX PS Claim 1; Page 1086-1096; 6221pp; English.
XX CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activity/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation.
XX CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX CC (AAM80020) are omitted as the relevant pages from the sequence listing
XX CC were missing at the time of publication.
XX SQ Sequence 7281 BP: 1831 A; 1773 C; 1915 G; 1762 T; 0 other;
Query Match 21.9%; Score 77.2; DB 22; Length 7281;
Best Local Similarity 86.7%; Pred. No. 7.5e-12;
Matches 85; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 255 AAACCCCGGTAAATTCGAGCGAGAGTGTAGTGGCGCGGGACCCGAGAGCGAGCGACCC 314
DB 22 AATTCCCGGGGTGCGAGGATTCGTGTGTGTGGCGCGGGACCCGAGAGCGAGCGACCC 81
QY 315 TTCTCTCCGGGCTGGCGAGGCGAGGGCGGGGAGCTC 352
DB 82 TTCTCTCCGGGCTGGCGAGGCGAGGGCGGGGAGCTC 119
Search completed: February 3, 2003, 16:33:17
Job time : 181.279 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 12:45:48 ; Search time 37.9628 Seconds
(without alignments)
8245.650 Million cell updates/sec

Title: US-09-596-141C-3 COPY 1394 1532

Perfect score: 139
Sequence: 1 ccccccgcgcctccagctct.....agcccaatcctataaaacaa 139

Scoring table: IDENTITY_NJC
Gapoc 10.0 Gapext 1.0

searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

```
Minimum D3 seq length: 0
Maximum D3 seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Genesec 101002:*

1:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA-995.DAT.*
17:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA-996.DAT.*
18:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SID82/cgcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result	Score	Query	Length	ID	Description
1	139	100.0	1197	24	Human large ATP-bi
2	139	100.0	1643	22	Nucleotide sequenc
3	139	100.0	1643	22	Nucleotide sequenc
4	139	100.0	183999	22	Human ABC1 genomic
5	137.4	98.8	2910	24	Human ABC1 transcr
6	137.4	98.8	3231	24	Human ABC1 transcr
7	36.4	26.2	38	22	ABC1 polymorphism
8	30	21.6	1817	19	Human RNA-binding
9	30	21.6	5868	22	Human digestive sv

ATTACHMENTS

RESULT. T 1

ABL58400 standard: DNA: 1197 BP.

AA
AC
ABI-58400-

XX 30-III.-2002 (first entry)

DI	30 000 2002 (first entry)	XX
DE	Human large ATP-binding cassette transporter 1 (hABCL) promoter sequence.	XX

Human: large ATP-binding cassette transporter 1; ABC1: promoter;

KW antitumor; large air binding cassette transporter 1; ABC1; ABCA1

xx
OS
Hom sapiens.

XX
1
2
3
4
5
6
7
8
9
10

PN WO200183506-

XX

PD 08-NOV-2001.

[illegible]

PF-21-APR-2
YY

28-APR-2000: 2000:JS-

XX
XX
PA (JYCO) UNIV COLUMBIA NEW YORK.
XX
PI Tall AR;
XX
DR wPI; 2002-049334/06.
XX
PT Novel isolated human large ATP-binding cassette transporter 1 promoter
PT capable of directing transcription of heterologous coding sequence
PT positioned downstream to it, useful for expressing foreign DNA in host
PT


```
FT /bound_moiety= "IK2"
FT 412..420
FT /*tag= g
FT /bound_moiety= "LYF1"
FT 528..539
FT /*tag= h
FT /bound_moiety= "LMO2COM/MYOD/DeltaEFl"
FT 549..556
FT /*tag= i
FT /bound_moiety= "LYF1"
FT 558..568
FT /*tag= j
FT /bound_moiety= "DeltaEFl"
FT 593..596
FT /*tag= k
FT /bound_moiety= "LXX2.5"
FT 608..620
FT /*tag= l
FT /bound_moiety= "NFY/CAAC"
FT 708..725
FT /*tag= m
FT /bound_moiety= "MZFl"
FT 723..730
FT /*tag= n
FT /bound_moiety= "MZFl"
FT 771..785
FT /*tag= o
FT /bound_moiety= "HFH2/SRY/EVIL"
FT 803..812
FT /*tag= p
FT /bound_moiety= "CREBp1/VBP"
FT 831..837
FT /*tag= r
FT /bound_moiety= "NKX2.5"
FT 1076..1089
FT /*tag= s
FT /bound_moiety= "GATA"
FT 1173..1188
FT /*tag= t
FT /bound_moiety= "LXRalpha/deltaEFl"
FT 1189..1209
FT /*tag= u
FT /bound_moiety= "DeltaEFl/LYF1/IK2"
FT 1483..1491
FT /*tag= v
FT /bound_moiety= "AP4"
FT 1498..1514
FT /*tag= w
FT /bound_moiety= "LMO2-COM/MYOD/deltaEFl/247"
FT 1524..1545
FT /*tag= x
FT /bound_moiety= "ZID/deltaEFl"
FT 1597..1607
FT /*tag= y
FT /bound_moiety= "DeltaEFl"
FT 1622..1627
FT /*tag= z
FT /bound_moiety= "PPAR"
FT 1632..1637
FT /*tag= aa
FT /bound_moiety= "PPAR"
FT 1685..1698
FT /*tag= ab
FT /bound_moiety= "USF/NMYC/MYCMAX"
FT 1787..1797
FT /*tag= ac
FT /bound_moiety= "DeltaEFl"
FT 1809..1819
FT /*tag= ad
FT /bound_moiety= "DeltaEFl"
FT 1822..1833
FT /*tag= ae
FT /bound_moiety= "SRY"
FT 1840..1850
FT /*tag= af
FT /bound_moiety= "AP1"
FT 1942..1956
FT /*tag= ag
FT /bound_moiety= "HNF3beta"
FT 1978..1985
FT /*tag= ah
FT /bound_moiety= "NKX2.5"
FT 2008..2016
FT /*tag= ai
FT /bound_moiety= "PPAR/NKX2.5/PPAR"
FT 2019..2024
FT /*tag= aj
FT /bound_moiety= "PPAR/NKX2.5/PPAR"
FT 2051..2059
FT /*tag= ak
FT /bound_moiety= "GATA"
FT 2104..2111
FT /*tag= al
FT /bound_moiety= "SOX5"
FT 2114..2152
FT /*tag= am
FT /bound_moiety= "SYR/HFH/HNF3beta"
FT 2221..2228
FT /*tag= an
FT /bound_moiety= "MZFl"
FT 2234..2249
FT /*tag= ao
FT /bound_moiety= "IK2/NFKappaB/CREL"
FT 2259..2272
FT /*tag= ap
FT /bound_moiety= "LMO2COM/GATA"
FT 2289..2306
FT /*tag= aq
FT /bound_moiety= "MZFl/SRY"
FT 2313..2318
FT /*tag= ar
FT /bound_moiety= "PPAR"
FT 2321..2326
FT /*tag= as
FT /bound_moiety= "PPAR"
FT 2335..2342
FT /*tag= at
FT /bound_moiety= "MZFl"
FT 2361..2384
FT /*tag= au
FT /bound_moiety= "HNF3beta/SRY/EVIL"
FT 2426..2433
FT /*tag= av
FT /bound_moiety= "MZFl"
FT 2442..2451
FT /*tag= aw
FT /bound_moiety= "AP4"
FT 2455..2466
FT /*tag= ax
FT /bound_moiety= "SRY"
FT 2491..2498
FT /*tag= ay
FT /bound_moiety= "STAT"
FT 2524..2534
FT /*tag= az
FT /bound_moiety= "STAT/PPAR"
FT 2536..2541
FT /*tag= ba
FT /bound_moiety= "PPAR"
FT 2589..2600
FT /*tag= bb
FT /bound_moiety= "AP2"
FT 2610..2617
FT /*tag= bc
FT /bound_moiety= "MZFl"
FT 2634..2648
```

```

FT      /*tag= bd
FI      /bound_moiety= "LMO2C04/WYOD/E47"
FI      2657..2672
FI      /tag= be
FI      /bound_moiety= "RREB1"
FT      2680..2698
FT      /tag= bf
FI      /bound_moiety= "MZFL/CMYB"
FT      2728..2740
FI      /tag= bg
FI      /bound_moiety= "SPL/GC"
FT      2743..2757
FI      /tag= bh
FI      /bound_moiety= "USF/NMYC/ARNT"
FT      2758..2773
FI      /tag= bi
FI      /bound_moiety= "NFE2AP1"
FT      2774..2787
FI      /tag= bj
FI      /bound_moiety= "XFD1/HFH"
FT      2794..2806
FI      /tag= bk
FI      /bound_moiety= "GC/SPL/MZF1"

```

Query Match 98.8%; Score 137.4; DB 24; Length 2910;
 Best Local Similarity 99.3%; Pred. No. 7.6e-35;
 Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY      1 GGGCCCCGGCTCCACGCTGCTTCTGCTGAGTGAAGTAACTACATACAGAGCCGGGAA 60
DB      |||||||
DB      2735 GGGCCCCGGCTCCACGCTGCTTCTGCTGAGTGAAGTAACTACATACAGAGCCGGGAA 2794

QY      61 GGGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
DB      |||||||
DB      2795 GGGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2854

QY      121 GCCGAATCTATAAAGGAA 139
DB      |||||||
DB      2855 GCCGAATCTATAAAGGAA 2873

```

```

RESULT 6
ID      AAD37265 standard; DNA; 3231 BP.
XX
AC      AAD37265;
XX
DT      21-AUG-2002 (first entry)
XX
DE      Human ABC1 transcription regulatory DNA #1.
XX
KW      Human; ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
KW      cholesterol metabolism; hypercholesterolaemia; antisense therapy; ds.
XX
OS      Homo sapiens.
XX
PN      WO200183746-A2.
XX
PD      08-NOV-2001.
XX
PF      02-MAY-2001; 2001WO-EP05488.
XX
PR      02-MAY-2000; 2000US-201280P.
XX
PA      (AVEI ) AVENTIS PHARMA SA.
XX
PI      Rosier-Montus M, Prades C, Lemoine C, Naudin L, Deneffie P;
PI      Brewer B, Duverger N, Remaley A, Santamarina-Fojo S;
XX
DR      WPI; 2002-154404/20.
XX
PT      Isolated nucleic acid useful for modifying the ATP-binding cassette 1
PT      (ABC1) and screening for candidate modulatory compounds or substances

```

```

PT
XX
PS      Claim 1: Page 130-131; 152pp; English.
XX
CC      The invention relates a nucleic acid which is capable of regulating the
CC      transcription of human ATP-binding cassette 1 (ABC1) gene, which is a
CC      casual gene for pathologies linked to a dysfunctioning of cholesterol
CC      metabolism, including diseases such as atherosclerosis. Polynucleotides
CC      of the invention are used to screen candidate molecules or substances
CC      that are capable of modulating the transcription of the ABC1 gene. They
CC      are used in antisense therapy. Compositions comprising sequences of the
CC      invention are used to treat hypercholesterolaemia and atherosclerosis.
CC      The present sequence is human ABC1 transcription regulating DNA.
XX
SQ      Sequence 3231 BP; 809 A; 773 C; 876 G; 773 T; 0 other;

```

Query Match 98.8%; Score 137.4; DB 24; Length 3231;
 Best Local Similarity 99.3%; Pred. No. 7.8e-35;
 Matches 138; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY      1 GGGCCCCGGCTCCACGCTGCTTCTGCTGAGTGAAGTAACTACATACAGAGCCGGGAA 60
DB      |||||||
DB      2735 GGGCCCCGGCTCCACGCTTCTGCTGAGTGAAGTAACTACATACAGAGCCGGGAA 2794

QY      61 GGGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
DB      |||||||
DB      2795 GGGGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2854

QY      121 GCCGAATCTATAAAGGAA 139
DB      |||||||
DB      2855 GCCGAATCTATAAAGGAA 2873

```

```

RESULT 7
ID      AAF93064 standard; DNA; 38 BP.
XX
AC      AAF93064;
XX
DT      17-MAY-2001 (first entry)
XX
DE      ABC1 polymorphism RFLP oligonucleotide #25.
XX
KW      High density lipoprotein-cholesterol; HDL-C; cardiovascular; ABC1; ds.
XX
OS      Homo sapiens.
XX
PN      WO200115676-A2.
XX
PD      08-MAR-2001.
XX
PF      01-SEP-2000; 2000WC-1B01492.
XX
PR      01-SEP-1999; 99US-0151977.
PR      15-MAR-2000; 2000US-0526193.
PR      23-JUN-2000; 2000US-0213958.
XX
PA      (UYBR-) UNIV BRITISH COLUMBIA.
PA      (XENO-) XENON GENETICS INC.
XX
PI      Hayden MR, Brooks-Wilson AR, Pi-stone SN, Clee SM;
XX
DR      WPI; 2001-244356/25.
XX
PT      Treating a lower than normal high density lipoprotein-cholesterol
PT      (HDL-C) level, a higher than normal triglyceride level, or a
PT      cardiovascular disease, by administering a compound that modulates LXR-
PT      or RXR-mediated transcriptional activity -
XX
PS      Disclosure; Fig 17; 317pp; English.
XX
CC      The present invention relates to a method for treating a patient
CC      diagnosed as having a lower than normal high density

```



```
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0233423.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234957.
PR 25-SEP-2000; 2000US-0234958.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240950.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241805.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 06-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246509.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 03-DEC-2000; 2000US-0251030.
PR 03-DEC-2000; 2000US-0251958.
PR 03-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251475.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0255678.

XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-502630/55.
XX
XX Polynucleotides encoding digestive system antigens, useful for
XX diagnosing, treating, preventing and/or prognosing disorders of the
XX digestive system, particularly cancer and cancer metastases -
XX
XX Disclosure: SEQ ID NO 2974; 986pp; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human digestive system antigens. These can be used in the
XX diagnosis, treatment and prevention of digestive system disorders,
XX including cancer, Meckel's diverticulum, bacterial or parasitic
XX infections, appendicitis, Hirschsprung's disease, chronic colitis or
XX ulcerative colitis. The present sequence is a genomic DNA fragment
XX encoding a digestive system antigen of the invention.
XX
XX Query Match 21.6%; Score 30; DB 22; Length 5868;
XX Best Local Similarity 61.5%; Pred. No. 9;
XX Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy 1 GGGCCCCGGCTCCAGCTGCTTTCGCTGAGTCAGTACATACATACAGAGAGCGCGGAA 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164 GGGCTGCTGCTCCTCGTCTCGGCGCTGATGGCGCGAAGACAGATATCAGGGCGAGCGCG 105
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 CGGGCGGGGAGGAGGA 78
   ||||| ||||| ||||| |||||
Db 104 CGGGCGGGGAGGAGGA 87
   ||||| ||||| ||||| |||||
```



```
RESULT 10
AAK89399/c
-D AAK89399 standard; DNA; 10316 BP.
XX
AC AAK89399;
XX
DI 05-NOV-2001 (first entry)
XX
XX Human digestive system antigen genomic sequence SEQ ID NO: 2975.
XX
XX Human; digestive system antigen; gene therapy; cancer; appendicitis;
XX ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
XX digestive system disorder; Meckel's diverticulum; ds.
XX
XX Homo sapiens.
XX
XX WC200155314-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-0561324.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209457.
XX 28-JUN-2000; 2000US-0214866.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220563.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 30-AUG-2000; 2000US-0227009.
XX 01-SEP-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
XX 08-SEP-2000; 2000US-0231244.
XX 08-SEP-2000; 2000US-0231413.
XX 08-SEP-2000; 2000US-0231414.
XX 08-SEP-2000; 2000US-0232083.
XX 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239035.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0248207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
```



```
FT CDS 1353..2674
FT /tag= a "RRV R1"
FT /product= "complement (2692...3259)"
FT CDS
FT /tag= b
FT /product= "dihydrofolate reductase"
FT /label= RRV_ORF2
FT /note= "has similarity to Kaposi's sarcoma-associated
FT virus (KSHV) open reading frame (ORF) 2"
FT CDS 3676..5613
FT /tag= c
FT /product= "complement binding protein"
FT /label= RRV_ORF3
FT /note= "has similarity to KSHV ORF4"
FT CDS 6045..9443
FT /tag= d
FT /product= "ssDNA binding protein"
FT /label= RRV_ORF6
FT /note= "has similarity to KSHV ORF6"
FT CDS 9468..11528
FT /tag= e
FT /product= "transport protein"
FT /label= RRV_ORF7
FT /note= "has similarity to KSHV ORF7"
FT CDS 11515..14004
FT /tag= f
FT /product= "glycoprotein B"
FT /label= RRV_ORF8
FT /note= "has similarity to KSHV ORF8"
FT CDS 14122..17166
FT /tag= g
FT /product= "DNA polymerase protein"
FT /label= RRV_ORF9
FT /note= "has similarity to KSHV ORF9"
FT CDS 17261..18511
FT /tag= h
FT /label= RRV_ORF10
FT /note= "has similarity to KSHV ORF10"
FT CDS 18520..19749
FT /tag= i
FT /label= RRV_ORF11
FT /note= "has similarity to KSHV ORF11"
FT CDS
FT /tag= j
FT /product= "RRV R2"
FT /label= RRV_ORF
FT /note= "has similarity to KSHV interleukin (IL)-6 gene"
FT CDS
FT /tag= k
FT /product= "thymidylate synthase"
FT /label= RRV_ORF70
FT /note= "has similarity to KSHV ORF70"
FT CDS
FT /tag= l
FT /product= "RRV R3"
FT /note= "has similarity to KSHV K4 viral MIP gene"
FT CDS 26846..27409
FT /tag= m
FT /product= "Bcl2-homologue"
FT /label= RRV_ORF16
FT /note= "has similarity to KSHV ORF16"
FT CDS
FT /tag= n
FT /label= RRV_ORF17
FT /note= "has similarity to KSHV ORF17"
FT CDS 28998..29897
FT /tag= o
FT /label= RRV_ORF18
FT /note= "has similarity to KSHV ORF18"
FT CDS
FT /tag= p
FT /product= "tegument protein"
FT /label= RRV_ORF19
FT /note= "has similarity to KSHV ORF19"
FT CDS
FT /note= "has similarity to KSHV ORF19"
FT complement (31043...32095)
FT /tag= q
FT /label= RRV_ORF20
FT /note= "has similarity to KSHV ORF20"
FT CDS 32094..33767
FT /tag= r
FT /product= "thymidine kinase"
FT /label= RRV_ORF21
FT /note= "has similarity to KSHV ORF21"
FT CDS 33754..35968
FT /tag= s
FT /product= "glycoprotein H"
FT /label= RRV_ORF22
FT /note= "has similarity to KSHV ORF22"
FT CDS
FT complement (35865...37073)
FT /tag= t
FT /label= RRV_ORF23
FT /note= "has similarity to KSHV ORF23"
FT CDS
FT complement (37123...39321)
FT /tag= u
FT /label= RRV_ORF24
FT /note= "has similarity to KSHV ORF24"
FT CDS 39323..43459
FT /tag= v
FT /product= "major capsid protein"
FT /label= RRV_ORF25
FT /note= "has similarity to KSHV ORF25"
FT CDS 43491..44408
FT /tag= w
FT /product= "capsid protein"
FT /label= RRV_ORF26
FT /note= "has similarity to KSHV ORF26"
FT CDS 44433..45242
FT /tag= x
FT /label= RRV_ORF27
FT /note= "has similarity to KSHV ORF27"
FT CDS 45408..45683
FT /tag= y
FT /label= RRV_ORF28
FT /note= "has similarity to KSHV ORF28"
FT CDS
FT complement (45733...46779)
FT /tag= z
FT /label= RRV_ORF29b
FT /note= "has similarity to KSHV ORF29b"
FT CDS 46905..47135
FT /tag= aa
FT /label= RRV_ORF30
FT /note= "has similarity to KSHV ORF30"
FT CDS 47093..47746
FT /tag= ab
FT /label= RRV_ORF31
FT /note= "has similarity to KSHV ORF31"
FT CDS 47883..49077
FT /tag= ac
FT /label= RRV_ORF32
FT /note= "has similarity to KSHV ORF32"
FT CDS 49049..50059
FT /tag= ad
FT /label= RRV_ORF33
FT /note= "has similarity to KSHV ORF33"
FT CDS
FT complement (49977...50960)
FT /tag= ae
FT /label= RRV_ORF29a
FT /note= "has similarity to KSHV ORF29a"
FT CDS 50959..51942
FT /tag= af
FT /label= RRV_ORF34
FT /note= "has similarity to KSHV ORF34"
FT CDS 51923..52372
FT /tag= ag
FT /label= RRV_ORF35
FT /note= "has similarity to KSHV ORF35"
```

```

FT CDS 52278..53585
FT /tag= ah
FT /product= "kinase"
FT /label= RVV_ORF36
FT /note= "has similarity to KSHV ORF36"
FT CDS 53566..55008
FT /tag= ai
FT /product= "alkaline exonuclease"
FT /label= RVV_ORF37
FT /note= "has similarity to KSHV ORF37"
FT CDS 54963..55172
FT /tag= aj
FT /label= RVV_ORF38
FT /note= "has similarity to KSHV ORF38"
FT complement (55255..56391)
FT /tag= ak
FT /product= "glycoprotein M"
FT /label= RVV_ORF39
FT /note= "has similarity to KSHV ORF39"
FT CDS 56526..57932
FT /tag= al
FT /product= "helicase/primase"
FT /label= RVV_ORF40
FT /note= "has similarity to KSHV ORF40"
FT CDS 57917..59528
FT /tag= am
FT /product= "helicase/primase"
FT /label= RVV_ORF41
FT /note= "has similarity to KSHV ORF41"
FT complement (58525..59343)
FT /tag= an
FT /label= RVV_ORF42
FT /note= "has similarity to KSHV ORF42"
FT complement (59297..61027)
FT /tag= ao
FT /product= "capsid protein"
FT /label= RVV_ORF43
FT /note= "has similarity to KSHV ORF43"

Query Match 20.7%; Score 28.8; DB 24; Length 128139;
Best Local Similarity 54.8%; Pred. No. 53;
Matches 57; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 10 CTCACGCTGTTTCIGCTGAGTGACTGAACATACATAACAGAGCGCGGGAACGGGGGG 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69172 CTTAACGAGTGTCTACTACTGCGGACCGAATAGCCATGCGCTGCCCCCGGAGTG 69113

QY 70 GAGGAGGGAGACAGAGCTTTTGACCGATAGTAACCTCTCGGCT 113
   || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69112 GATATGGAGATGTGAAGCAGGAGAGATGTGCACCCGCCCT 69069

RESULT 15
AAC64754/c
ID AAC64754 standard; DNA; 133719 BP.
XX AC
XX AAC64754;
XX AC
XX 28-FEB-2001 (first entry)
XX DE
XX Macaca mulatta rhadinovirus 17577 (RRV) genome sequence SEQ ID NO:1.
XX
XX Macaca mulatta rhadinovirus 17577; RRV; rhesus macaque rhadinovirus;
XX genome; Kaposi's sarcoma-associated herpesvirus; KSHV; interleukin 6;
XX IL-6; macrophage inflammatory protein; MIP; diagnosis; vaccine;
XX cytostatic; anti-HIV; gene therapy; infection; Kaposi's sarcoma;
XX lymphoproliferative disorder; B-cell hyperplasia; lymphadenopathy;
XX splenomegaly; hypergammaglobulinemia; autoimmune haemolytic anaemia;
XX ds.
XX
XX Macaca mulatta rhadinovirus 17577.
XX OS
XX WO200028040-A2.
PN

```

```

XX 18-MAY-2000.
XX
XX 05-NOV-1999; 99WO-US26260.
XX
XX 06-NOV-1998; 98US-0107507.
XX 20-NOV-1998; 98US-0109409.
XX
XX (UYOR-) UNIV OREGON HEALTH SCI.
XX
XX Wong SW, Axthelm MK, Searles RP;
XX WPI: 2000-376552/32.
XX
XX New rhesus rhadino virus for producing non-human primate useful
XX for testing potential treatments and efficacy of the candidate vaccine
XX for conditions associated with RRV infection -
XX
XX Claim 2; Page 83-122; 141pp; English.
XX
XX The present invention describes a novel rhesus macaque rhadinovirus
XX called macaca mulatta rhadinovirus 17577 (RRV). AAC64754 represents the
XX RRV genome sequence, and AAB53123 to AAB53204 represent the proteins
XX encoded by the genome sequence. The present invention also specifically
XX claims the individual open reading frame (ORF) nucleotide sequences from
XX the genome which encode the individual proteins, but these sequences are
XX not given. A non-human animal infected with RRV can be used for testing
XX the efficacy of drug in the treatment of condition associated with
XX infection with RRV such as Kaposi's sarcoma, lymphoproliferative
XX disorders, B-cell hyperplasia, lymphadenopathy, splenomegaly,
XX hypergammaglobulinemia or autoimmune haemolytic anaemia, by
XX administering the drug to a immuno-compromised non-human primate
XX preferably Rhesus macaque monkey obtained by as a result of infection
XX by Simian Immunodeficiency Virus (SIV). RRV is useful for producing
XX non-human primate model for testing potential treatments for conditions
XX associated with RRV infection. It is also useful for testing the
XX efficacy of the candidate vaccine against RRV infection or conditions
XX associated with its infection by administering the vaccine to the
XX subject capable of infection with RRV, inoculating the subject with RRV
XX and observing the effect of vaccine. AAC64755 to AAC64765 and AAB53205
XX to AAB53213 represent sequence used in the exemplification of the
XX present invention.
XX
XX SQ Sequence 133719 BP; 32746 A; 35648 C; 34521 G; 30804 I; 0 other;
XX
XX Query Match 20.7%; Score 28.8; DB 21; Length 133719;
XX Best Local Similarity 54.8%; Pred. No. 54;
XX Matches 57; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
XX
XX 10 CTCACGCTGTTTCIGCTGAGTGACTGAACATACATAACAGAGCGCGGGAACGGGGGG 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 74752 CTTAACGAGTGTCTACTACTGCGGACCGAATAGCCATGCGCTGCCCCCGGAGTG 74693

QY 70 GAGGAGGGAGACAGAGCTTTTGACCGATAGTAACCTCTCGGCT 113
   || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 74692 GATATGGAGATGTGAAGCAGGAGAGATGTGCACCCGCCCT 74649

Search completed: February 3, 2003, 16:39:33
Job time : 317.106 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 14:33:17 ; Search time 31.8341 Seconds
(without alignments)
7959.555 Million cell updates/sec

Title: US-09-596-141C-3_COPY_1080_1643

Perfect score: 564
Sequence: 1 GAGCTCTCTCTCCCAATC.....CAGGCAAGGCGGAGGCTC 564

Scoring table: IDENTITY_NGC

Gapop 10.0 , Gapext 1.0

Searched: 396772 seqs, 224632407 residues

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications, NA: *

1: /cgn2_6/ptodata/1/pubnpa/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubnpa/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubnpa/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubnpa/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubnpa/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubnpa/PCTCS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubnpa/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubnpa/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubnpa/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubnpa/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubnpa/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubnpa/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubnpa/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubnpa/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	547.4	97.1	3231	10	US-09-846-456-1
2	456.4	80.9	2893	10	US-09-846-456-3
3	91	16.1	221	10	US-09-846-456-4
4	38	6.7	98665	10	US-09-770-889A-3
5	37.4	5.6	42450	10	US-09-815-048-3
6	37.2	5.6	1920	9	US-10-047-542-44
7	37.2	5.6	42999	9	US-09-799-462A-17
8	37.2	5.6	42999	9	US-10-125-767-17
9	36.6	6.5	4826	10	US-09-772-304A-1
10	36.2	6.4	1102	10	US-09-833-381-1385
11	35.6	6.3	1797	10	US-09-834-975-987
12	35.6	6.3	1797	10	US-09-834-975-1003
13	35.6	6.3	1797	10	US-09-834-975-1012
14	35.6	6.3	1797	10	US-09-834-975-1024
15	34.6	6.1	15500	10	US-09-764-869-2109
16	34.4	6.1	252	10	US-09-923-876-250
17	34.4	6.1	1699	10	US-09-819-254-1
18	34.4	6.1	1699	10	US-09-880-107-3381
19	34.4	6.1	1699	10	US-09-998-596-2587

C	20	34.4	6.1	1856	10	US-09-895-652-3	Sequence 3, Appli
	21	34.4	6.1	1984	9	US-09-834-133-390	Sequence 390, App
	22	34.4	6.1	1984	10	US-09-738-973-390	Sequence 390, App
C	23	34.2	6.1	1490	9	US-10-000-273-3	Sequence 3, Appli
C	24	34.2	6.1	2422	10	US-09-833-381-1993	Sequence 1993, Ap
C	25	34.2	6.1	2460	9	US-10-000-273-1	Sequence 1, Appli
C	26	33.8	6.0	3715	9	US-09-965-830-9	Sequence 9, Appli
C	27	33.8	6.0	42999	9	US-09-799-462A-17	Sequence 17, Appli
C	28	33.8	6.0	42999	9	US-10-125-767-17	Sequence 17, Appli
	29	33.6	6.0	2091	9	US-10-199-333-1	Sequence 1, Appli
	30	33.6	6.0	5025	10	US-09-960-253-176	Sequence 176, App
	31	33.4	5.9	363	10	US-09-783-590-6604	Sequence 6604, Ap
C	32	33.4	5.9	4723	9	US-09-373-967-3	Sequence 3, Appli
C	33	33.4	5.9	5025	10	US-09-960-253-176	Sequence 176, App
C	34	33.4	5.9	43058	10	US-09-954-456-292	Sequence 292, App
C	35	33.4	5.9	43058	10	US-09-954-456-529	Sequence 529, App
C	36	33.4	5.9	43058	10	US-09-880-107-3950	Sequence 3950, Ap
C	37	33.2	5.9	1157	9	US-10-125-470-6	Sequence 6, Appli
C	38	33.2	5.9	1157	9	US-10-125-452-6	Sequence 6, Appli
C	39	33.2	5.9	1157	9	US-09-955-504-6	Sequence 6, Appli
	40	33.2	5.9	2540	10	US-09-917-800A-1636	Sequence 1636, Ap
	41	33.2	5.9	3236	10	US-09-954-456-43	Sequence 43, Appli
	42	33.2	5.9	3750	9	US-09-964-899-32	Sequence 32, Appli
	43	33.2	5.9	4071	12	US-10-044-090-591	Sequence 591, App
	44	32.8	5.8	860	10	US-09-764-870-149	Sequence 149, App
C	45	32.8	5.8	4286	10	US-09-899-634A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-846-456-1
; Sequence 1, Application US/09846456
; Patent No. US20020146792A1
; GENERAL INFORMATION:
; APPLICANT: Rosier, Marie
; APPLICANT: Prades, Catherine
; APPLICANT: Lemoine, Cendrine
; APPLICANT: Naudin, Laurent
; APPLICANT: Deneffe, Patrice
; APPLICANT: Duverger, Nicolas
; APPLICANT: Brewer, Bryan
; APPLICANT: Remaley, Alan
; APPLICANT: FC30, Silvia
; TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying
; FILE OF INVENTION: Activity and Therapeutic Uses
; FILE REFERENCE: 3806.0505
; CURRENT APPLICATION NUMBER: US/09/846,456
; CURRENT FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 50/201,280
; PRIOR FILING DATE: 2000-05-02
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 3231
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-846-456-1

QY	1	GAGCTCTCTCTCCCAATCCTCCCTCCGCTGAGGAACCTAACAGGAAAAAAT	60	Sequence 3, Appli
DB	2416	GAGCTCTCTCTCCCAATCCTCCCTCCGCTGAGGAACCTAACAGGAAAAAAT	2475	Sequence 3, Appli
QY	61	TGCGGAACGAGATTACAGAGCAAAATTCACCTGGTGGCTGGCGGAGCT	120	Sequence 1, Appli
DB	2476	TGCGGAACGAGATTACAGAGCAAAATTCACCTGGTGGCTGGCGGAGCT	2535	Sequence 1, Appli
QY	121	GGACTAGAGAGTCTCGGGCGGAGCCCGAGCCAGGCTTCGGCCCTTTAGGCCCGC	180	Sequence 1, Appli

Db 2536 GAGCTAGAGAGTCTGGCGGCGAGCCCGAGCCGCTTCCGCGCTTATAGGCCGC 2595
QY 181 GGGCCCGGGGGGAGGGGAGGAGAGCGGGGAGCCCTAAGACACCTGCTGTACCTCC 240
Db 2596 GGGCCCGGGGGGAGGGGAGGAGAGCGGGGAGCCCTAAGACACCTGCTGTACCTCC 2655
QY 241 AC-----CCCGACCCACCCACCTCCCGCCAACTCCCTAGATGTGTGCGGGGGGTGAA 295
Db 2655 ACCCCACCCACCCACCCACCTCCCGCCAACTCCCTAGATGTGTGCGGGGGGTGAA 2715
QY 296 CGTGGCCCGCTTAAAGGGGGGGGCGGCGGCTCCACGCTCTTCTGCTGAGTGAACCTA 355
Db 2716 CGTGGCCCGCTTAAAGGGGGGGGCGGCGGCTCCACGCTCTTCTGCTGAGTGAACCTA 2775
QY 356 CATAAACAGAGCGCGGAGCGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 415
Db 2776 CATAAACAGAGCGCGGAGCGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 2835
QY 416 AACCTCTGGCTCGGTGCAGCCGAATCTATAAAGGAAGTGTCCCGGCAAAACCCGT 475
Db 2836 AACCTCTGGCTCGGTGCAGCCGAATCTATAAAGGAAGTGTCCCGGCAAAACCCGT 2895
QY 476 AATGGAGGAGAGTGTAGTGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGTCTCC 535
Db 2896 AATGGAGGAGAGTGTAGTGGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGTCTCC 2955
QY 536 GGGCTGGCGAGGCGAGCGGGGGGAGGAGTCT 564
Db 2956 GGGCTGGCGAGGCGAGCGGGGGGAGGAGTCT 2984

RESULT 2
US-09-846-456-3
; Sequence 3, Application US/09846456
; Patent No. US20020146792A1
; GENERAL INFORMATION:
; APPLICANT: Rosier, Marie
; APPLICANT: Prades, Catherine
; APPLICANT: Lemoine, Cendrline
; APPLICANT: Naudin, Laurent
; APPLICANT: Denefle, Patrice
; APPLICANT: Duverger, Nicolas
; APPLICANT: Brewer, Bryan
; APPLICANT: Remaley, Alan
; APPLICANT: Fojo, Silvia
; TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying It
; FILE REFERENCE: 3806.0505
; CURRENT APPLICATION NUMBER: US/09/846,456
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/201,280
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent version 3.0
; SEQ ID NO 3
; LENGTH: 2893
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-846-456-3

Query Match 80.9%; Score 456.4; DB 10; Length 2893;
Best Local Similarity 98.7%; Pred. No. 2.1e-115;
Matches 472; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY 1 GAGCTCTCTCCCGCAATCCCTCCCTCCGCTGTGAGGAAGCAATTCAGTGGTCCCTTGGCTGCCGGGAACGT 60
Db 2416 GAGCTCTCTCCCGCAATCCCTCCCTCCGCTGTGAGGAAGCAATTCAGTGGTCCCTTGGCTGCCGGGAACGT 2475
QY 61 TGGCAAGCAGGATTATAGAGGAGCAAAATTCAGTGGTCCCTTGGCTGCCGGGAACGT 120
Db 2476 TGGCAAGCAGGATTATAGAGGAGCAAAATTCAGTGGTCCCTTGGCTGCCGGGAACGT 2535

QY 121 GAGCTAGAGAGTCTGGCGGCGAGCCCGAGCCGCTTCCGCGCTTATAGGCCGC 180
Db 2536 GAGCTAGAGAGTCTGGCGGCGAGCCCGAGCCGCTTCCGCGCTTATAGGCCGC 2595
QY 181 GGGCCCGGGGGGAGGGGAGGAGAGCGGGGAGCCCTAAGACACCTGCTGTACCTCC 240
Db 2596 GGGCCCGGGGGGAGGGGAGGAGAGCGGGGAGCCCTAAGACACCTGCTGTACCTCC 2655
QY 241 AC-----CCCGACCCACCCACCTCCCGCCAACTCCCTAGATGTGTGCGGGGGGTGAA 295
Db 2655 ACCCCACCCACCCACCCACCTCCCGCCAACTCCCTAGATGTGTGCGGGGGGTGAA 2715
QY 296 CGTGGCCCGCTTAAAGGGGGGGGCGGCGGCTCCACGCTCTTCTGCTGAGTGAACCTA 355
Db 2716 CGTGGCCCGCTTAAAGGGGGGGGCGGCGGCTCCACGCTCTTCTGCTGAGTGAACCTA 2775
QY 356 CATAAACAGAGCGCGGAGCGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 415
Db 2776 CATAAACAGAGCGCGGAGCGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 2835
QY 416 AACCTCTGGCTCGGTGCAGCCGAATCTATAAAGGAAGTGTCCCGGCAAAACCCGT 473
Db 2836 AACCTCTGGCTCGGTGCAGCCGAATCTATAAAGGAAGTGTCCCGGCAAAACCCGT 2893
RESULT 3
US-09-846-456-4
; Sequence 4, Application US/09846456
; Patent No. US20020146792A1
; GENERAL INFORMATION:
; APPLICANT: Rosier, Marie
; APPLICANT: Prades, Catherine
; APPLICANT: Lemoine, Cendrline
; APPLICANT: Naudin, Laurent
; APPLICANT: Denefle, Patrice
; APPLICANT: Duverger, Nicolas
; APPLICANT: Brewer, Bryan
; APPLICANT: Remaley, Alan
; APPLICANT: Fojo, Silvia
; TITLE OF INVENTION: Regulatory Nucleic Acid for the ABC1 Gene, Molecules Modifying
; FILE REFERENCE: 3806.0505
; CURRENT APPLICATION NUMBER: US/09/846,456
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/201,280
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent version 3.0
; SEQ ID NO 4
; LENGTH: 221
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-846-456-4

Query Match 16.1%; Score 91; DB 10; Length 221;
Best Local Similarity 100.0%; Pred. No. 6.1e-16;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 474 GTAAATTCGAGCGAGAGTGTGGGCGGCGGAGCCCGAGAGCCGAGAGCCGAGCCGCTTCTCTC 533
Db 1 GTAAATTCGAGCGAGAGTGTGGGCGGCGGAGCCCGAGAGCCGAGAGCCGAGCCGCTTCTCTC 60
QY 534 CGGGGTGTGGCGAGGCGAGGCGGGGAGGAGTCT 564
Db 61 CGGGGTGTGGCGAGGCGAGGCGGGGAGGAGTCT 91

RESULT 4
US-09-770-689A-3
; Sequence 3, Application US/09770689A
; Patent No. US2002011517A1
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.


```

; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001079
; CURRENT APPLICATION NUMBER: US/09/770,689A
; CURRENT FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 98665
; TYPE: DNA
; ORGANISM: HUMAN
US-09-770-689A-3

Query Match          6.7%;   Score 38;   DB 10;   Length 98665;
Best Local Similarity 55.5%;   Pred. No. 1.7;
Matches 74;   Conservative 0;   Mismatches 60;   Indels 0;   Gaps 0;

QY 74 ATTGAGGGAAGCAAAATTCACCTGGTCCCTTGGCTCCCGGGAACGTGGACTAGAGAGATC 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2100 ATGAAGAGCAGTTCACCGCATCAACAGCTGGCTAACAGACCGCTGGCAGCGGAGIG 2159

QY 134 TGGCGCCAGCCCGCGAGCCAGCGCTTCCCGCGCGGCTTTAGCGCGGCGGCGCCCGGGCGGG 193
    || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db 2160 CGCGGGCAGCAGCGGGGCTCGACCGGGGCTTGGGGCGGAGCGCGGAGCGCGGGCGC 2219

QY 194 GGAAGGGGACGGCAG 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2220 GGACGGCTCCTCCG 2233

RESULT 5
US-09-815-048-3
; Sequence 3, Application US/09815048
; Patent No. US2002013713A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLG01180
; CURRENT APPLICATION NUMBER: US/09/815,048
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 42450
; TYPE: DNA
; ORGANISM: Human
US-09-815-048-3

Query Match          6.5%;   Score 37.4;   DB 10;   Length 42450;
Best Local Similarity 51.5%;   Pred. No. 1.9;
Matches 86;   Conservative 0;   Mismatches 81;   Indels 0;   Gaps 0;

QY 106 GGCTGCGCGGAACGTGAGCTAGAGAGTGTGGCGCGACCCCGAGCCCGAGCGCTTCCCGC 165
    ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1537 GGCTCCCGGACGCAAGCCTTAGAGGCGCGGAGAGGCGCCGCCGCCGCCCTTCGGCC 1596

QY 166 CGCTCTTAGCGCGCGCGCGCGGGGGAAGGGAGCGACCGAGCCGCACTTAAGACA 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1597 CCACCCACCAACCCCGCGCCCGCACCGACCGACCGAGGCGCCGCCGCCCTCCCAACCA 1656

QY 226 CTGCTGTAGCTTCGACCCCGCACCCACCGACCTCCCGCCCACTCC 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1657 CCAGCGCGCGCCCACTTCGACCCACCGACCGAGCGCGCGCTCATGCC 1703

RESULT 6
US-10-047-542-44
; Sequence 44, Application US/10047542
; Patent No. US2002016837A
; GENERAL INFORMATION:

```

```

; APPLICANT: LARRICK, JAMES W.
; APPLICANT: MYCOFF, KEITH L.
; TITLE OF INVENTION: NOVEL IMMUNOADHESINS FOR TREATING AND PREVENTING VIRAL
; TITLE OF INVENTION: AND BACTERIAL DISEASES
; FILE REFERENCE: 030905.0034.C1P1
; CURRENT APPLICATION NUMBER: US/10/047.542
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: PCT/US01/13932
; PRIOR FILING DATE: 2001-04-28
; PRIOR APPLICATION NUMBER: 60/200,298
; PRIOR FILING DATE: 2000-04-28
; NUMBER OF SEQ ID NOS: 101
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-047-542-44

      Query Match      6.6%; Score 37.2; DB 9; Length 1920:
      Best Local Similarity 49.5%; Pred. No. 0.67;
      Matches 96; Conservative 0; Mismatches 98; Indels 0; Gaps 0

QY 148 GAGCCACAGCCCTCCGCGCGGTCTTAGCCGCGCGCGCGCGGGAAGGGAGCGCAG 207
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 508 GAACACCCAGCATGCTCGGACCGCGGCGAGGACGTGGGGCGAAGAGGGACACAG 567
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 208 ACGCGGAGCCCTAAGACACCTGCTGTACCCCTCCACCCCGCCACCCACCTCCCGCCAA 267
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 568 AGCTAGAGAGCGGCACACCCCTCATGACCACCGACTTCCCCAGTCTGCTCCAGGGAC 627
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 268 CTCCTTAGATGTCGTGGCGGCTGAACGTCGCGCCCTTAAAGGGCGCGCGCGCTCC 327
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 628 TTCACCCCGCCACCGTGAAGATCTTACAGTCGTCTCGACGCGCGCGGCGACTTCCCG 687
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 328 ACGTGCTTCTGCT 341
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 688 CCGACCATCCAGCT 701
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
US-09-799-462A-17
; Sequence 17, Application US/09799462A
; Patent No. US2003016097CA.
; GENERAL INFORMATION:
; APPLICANT: Hadlaczky, Gyula
; Szalay, Aladar
; TITLE OF INVENTION: ARTIFICIAL CHROMOSOMES, USES THEREOF
; AND METHODS FOR PREPARING ARTIFICIAL CHROMOSOMES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/799,462A
; FILING DATE: 10-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/835,682
; FILING DATE: 10-APR-1997
; APPLICATION NUMBER: 08/695,191
; FILING DATE: 07-AUG-1996
; APPLICATION NUMBER: 08/682,080
; FILING DATE: 15-JUL-1996

```


QY	203	CGCAGACGGGGACCTTAAGACACCTGCTGTACCTTCACGCCACGCCACCTCC	262
Db	887	ACCCAGCGGGACCGGAGCCCTCGGGTCCCCGCCACGTCGGACAGCACGTCAC	828
QY	263	CCCACTCCCTAGATGTGCTGGCGGGCTGAAGCTGGCCCGTTTAAGGGGGGGCCCG	322
Db	827	CGGGCCACAGACACGGCGGAGGGGGAGCGCCACACAGCGGGCGCGCACCA	758
QY	323	GCTCCAC	329
Db	767	CCGCCCC	761

```

RESULT 10
US-09-833-381-1386
; Sequence 1386, Application US/09833381
; Patent NO. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: NO. US20020132090A1el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1386
; LENGTH: 1102
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1102)
; OTHER INFORMATION: n = A,T,C or G
US-09-833-381-1386

```

APPLICANT: Lallie, James
 APPLICANT: Brown, Jeffrey
 APPLICANT: Bolt, Andrew
 APPLICANT: Van Huffel, Christophe
 TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
 FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
 OF HUMAN CANCERS
 FILE REFERENCE: MRI-016B
 CURRENT APPLICATION NUMBER: US/09/834,975
 CURRENT FILING DATE: 2001-04-13

Search completed: February 3, 2003, 16:30:02
Job time : 130.834 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 16:38:05 ; Search time 4386.17 Seconds
(without alignments)
10165.006 Million cell updates/sec

Title: US-09-596-141C-3_COPY_1_1532

Perfect score: 1532

Sequence: 1 gaattccttggtggtgc.....agccgaatctataaaagggaa 1532

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2034640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Lasting first 45 summaries

Database :

GccEmbl.*

1: gb_ba.*

2: gb_hg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vl.*

30: em_hg_hum.*

31: em_hg_inv.*

32: em_hg_other.*

33: em_hg_mus.*

34: em_hg_pln.*

35: em_hg_rtd.*

36: em_hg_nam.*

37: em_hg_vrt.*

38: em_sy.*

39: em_hgo_hum.*

40: em_hgo_mus.*

41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1532	100.0	1643	6	AX060715	Sequence
2	1532	100.0	1643	6	AX060894	Sequence
3	1433.8	93.6	96717	9	AL359182	Human DNA
4	1387.6	90.6	183999	6	AX092589	Sequence
5	1374.4	89.7	2893	6	AX351031	Sequence
6	1374.4	89.7	3231	6	AX351029	Sequence
7	1307	85.3	149034	9	AF275948	Sequence
8	1138.2	74.3	201144	9	AF275948	Homo sapi
9	896	58.5	175064	2	AC012230	Homo sapi
10	896	58.4	1157	9	HS4252201	Homo sapi
11	893.4	58.3	1167	9	AF258623S1	Homo sapi
12	408.8	26.7	69370	2	AC021246	Homo sapi
13	171.8	11.2	69570	2	AC021246	Homo sapi
14	135.6	8.9	90688	2	AC021345	Homo sapi
15	128.8	8.4	186889	2	AL807243	Mus muscu
16	128.8	8.4	278572	10	AF287263	Mus muscu
17	126.6	8.3	145833	2	AC125837	Rattus no
18	102.2	6.7	2893	6	AX351031	Sequence
19	102.2	6.7	3231	6	AX351029	Sequence
20	99.4	6.5	96717	9	AL359182	Sequence
21	97.8	6.4	49034	9	AF275948	Homo sapi
22	97	6.3	1643	6	AX060715	Sequence
23	97	6.3	1643	6	AX060894	Sequence
24	92.6	6.0	183999	6	AX092589	Sequence
25	69.2	4.5	201144	9	AF287262	Homo sapi
26	62	4.0	897	9	AF258627	Homo sapi
27	55.6	3.6	191720	2	AC109982	Rattus no
28	52.6	3.4	125020	9	AF429315	Homo sapi
29	52.2	3.4	125020	9	AF429315	Homo sapi
30	49.6	3.2	96136	2	AC097674	Rattus no
31	49.4	3.2	173127	2	AC131142	Rattus no
32	49	3.2	65780	2	AC111741	Rattus no
33	48.6	3.2	13645	2	AC108454	Homo sapi
34	48.4	3.2	137136	2	AC108905	Homo sapi
35	48.4	3.2	141114	2	AC108903	Felis cat
36	48.4	3.2	206789	2	AC109180	Mus muscu
37	47.6	3.1	151975	2	AC109403	Rattus no
38	47.6	3.1	190343	2	AC097794	Rattus no
39	47.4	3.1	101492	2	AC116065	Rattus no
40	47.4	3.1	180461	2	AC106565	Rattus no
41	47	3.1	106983	9	AC108077	Homo sapi
42	47	3.1	110000	2	AC096324_0	Rattus no
43	47	3.1	167886	2	AC126316	Rattus no
44	46.8	3.1	2870	9	HS4800882	Homo sapi
45	46.8	3.1	169799	2	AC098129	Rattus no

ALIGNMENTS

RESULT 1
AX060715
LOCUS AX060715
DEFINITION Sequence 3 from Patent WO0078972.
ACCESSION AX060715
VERSION AX060715.1 GI:12406104
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1643)
AUTHORS Lawn, R.M., Wade, D. and Garvin, M.
TITLE Regulation with binding cassette transporter protein abcl
JOURNAL Patent: WO 0078972-A 3 28-DEC-2000;

```
CV THERAPEUTICS, INC. (US)
Location/Qualifiers
1..1643
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 370 a 413 c 457 g 403 t
ORIGIN
Query Match 100.0%; Score 1532; DB 6; Length 1643;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCCTGCTGGTCCACATGCACTTCCAGGCGCTGCTGGCTCTCTAATGGGTC 60
Db 1 GAATTCCTGCTGGTCCACATGCACTTCCAGGCGCTGCTGGCTCTCTAATGGGTC 60
QY 61 TGTCTGAGTCTTGAAGAACCACTGATGTCAGTACCTGGCTGAGCGCTGGGCTGGAGA 120
Db 61 TGTCTGAGTCTTGAAGAACCACTGATGTCAGTACCTGGCTGAGCGCTGGGCTGGAGA 120
QY 121 TCCGTGTTGACTGTAGCATGAGGGGCTTGTACGCTGAATGCTGTATGCAAGTGGTGGG 180
Db 121 TCCGTGTTGACTGTAGCATGAGGGGCTTGTACGCTGAATGCTGTATGCAAGTGGTGGG 180
QY 181 ACTTCTGGAATATGATGAGCTGGAGGTGGGAAGAGAAGTAGGCTTGGGCGAGCTCTC 240
Db 181 ACTTCTGGAATATGATGAGCTGGAGGTGGGAAGAGAAGTAGGCTTGGGCGAGCTCTC 240
QY 241 ATGCCACCTCTATTCGGCCAAACTCAGCTCAAACTGTGAAGAGTCTAAATGTGAATCTG 300
Db 241 ATGCCACCTCTATTCGGCCAAACTCAGCTCAAACTGTGAAGAGTCTAAATGTGAATCTG 300
QY 301 CCTTCAAGTGGGTACAAAGTATCTTTGTCAAGTAGGAGACCTTGTGGCCCTCCAGCT 360
Db 301 CCTTCAAGTGGGTACAAAGTATCTTTGTCAAGTAGGAGACCTTGTGGCCCTCCAGCT 360
QY 361 GCATTCGAGGCGCTGCTGGCTCTTCAAGGCTCTGCTCTGCTGCTGCTCTGCTCTGCT 420
Db 361 GCATTCGAGGCGCTGCTGGCTCTTCAAGGCTCTGCTCTGCTGCTGCTCTGCTCTGCT 420
QY 421 CCTTCAGGCGAGATCATATTTAGACTCTTACAGCTCTTACAGCTTGTGAGTCTGAGTCT 480
Db 421 CCTTCAGGCGAGATCATATTTAGACTCTTACAGCTCTTACAGCTTGTGAGTCTGAGTCT 480
QY 481 AGGTGACATTTAGTTTGTGCTCTGATGAATGACTTAAATATTTAGACATATGTTGTGTA 540
Db 481 AGGTGACATTTAGTTTGTGCTCTGATGAATGACTTAAATATTTAGACATATGTTGTGTA 540
QY 541 GGCCTGCACTTCTACTCTGCTCTTTTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 600
Db 541 GGCCTGCACTTCTACTCTGCTCTTTTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 600
QY 601 CCCCTACAGCCAAAGGCAACAGATAAGTGGAGGCTGGAGGCTGGCTACATAATTTTAC 660
Db 601 CCCCTACAGCCAAAGGCAACAGATAAGTGGAGGCTGGAGGCTGGCTACATAATTTTAC 660
QY 661 ACAGCTGCAATCTCTGCTGCACTTCAAAAGTATACAAACATAAATACAAAGTCTCTG 720
Db 661 ACAGCTGCAATCTCTGCTGCACTTCAAAAGTATACAAACATAAATACAAAGTCTCTG 720
QY 721 TTTTATCACAGGAGGCTGATCAATATAATGAAATTAAGAGGGGCTGGTCCCATATTTG 780
Db 721 TTTTATCACAGGAGGCTGATCAATATAATGAAATTAAGAGGGGCTGGTCCCATATTTG 780
QY 781 TCTGCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 840
Db 781 TCTGCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 840
QY 841 ATGAAGAGAGCAGTAAAGATGTTCTCTGGGCTCTCTGAGGAGCTGGGAGCTCAGGC 900
Db 841 ATGAAGAGAGCAGTAAAGATGTTCTCTGGGCTCTCTGAGGAGCTGGGAGCTCAGGC 900
QY 901 TGGGAATCTCCAGGCACTAGTCCCTATCAAAATCAAAATCAAAATCAAAATCAAAATCA 960
Db 901 TGGGAATCTCCAGGCACTAGTCCCTATCAAAATCAAAATCAAAATCAAAATCAAAATCA 960
QY 961 AAACAAAAGCAGCCCATTTACCCAGAGACTGTCGGCTTCCCTTCCCTTCCCTTCCCTT 1020
Db 961 AAACAAAAGCAGCCCATTTACCCAGAGACTGTCGGCTTCCCTTCCCTTCCCTTCCCTT 1020
QY 1021 TTTGAAAGGAAACAAAGACAAAGCAAAATGATTGGCTCTCTGAGGAGATTTCAGCCTAG 1080
Db 1021 TTTGAAAGGAAACAAAGACAAAGCAAAATGATTGGCTCTCTGAGGAGATTTCAGCCTAG 1080
QY 1081 AGCTCTCTCTCCCCAATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1140
Db 1081 AGCTCTCTCTCCCCAATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1140
QY 1141 GCGAAAGCAGAGATTAGAGGAAGCAAAATCCACITGTCCTTGGCTTGGCTTGGCTTGG 1200
Db 1141 GCGAAAGCAGAGATTAGAGGAAGCAAAATCCACITGTCCTTGGCTTGGCTTGGCTTGG 1200
QY 1201 GACTAGAGAGTCTGGGCGCAGCCCGGAGCCCGCTTCCCGCGCGCTTTCAGCCGCGG 1260
Db 1201 GACTAGAGAGTCTGGGCGCAGCCCGGAGCCCGCTTCCCGCGCGCTTTCAGCCGCGG 1260
QY 1261 GCGCGGCGGCGGGAAGGCGAGCCCGGAGCCCGCTTCCCGCGCGCTTTCAGCCGCGG 1320
Db 1261 GCGCGGCGGCGGGAAGGCGAGCCCGGAGCCCGCTTCCCGCGCGCTTTCAGCCGCGG 1320
QY 1321 CCCCCACCCACCCACCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1380
Db 1321 CCCCCACCCACCCACCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1380
QY 1381 CCGTTTAAAGGCGGCGGCGGCTCCAGCTCTTCTGCTGAGTCACTGAATACATAAAA 1440
Db 1381 CCGTTTAAAGGCGGCGGCGGCTCCAGCTCTTCTGCTGAGTCACTGAATACATAAAA 1440
QY 1441 CAGAGCGCGGGAAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1500
Db 1441 CAGAGCGCGGGAAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
QY 1501 TGCGTCTGCTGCGAGCGCAATCTATAAAAGGAA 1532
Db 1501 TGCGTCTGCTGCGAGCGCAATCTATAAAAGGAA 1532
RESULT 2
AX060894
LOCUS AX060894 1643 bp DNA linear PAT 22-JAN-2001
DEFINITION Sequence 3 from Patent WO0078971.
ACCESSION AX060894
VERSION AX060894.1 GI:12406271
KEYWORDS
SOURCE Human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 1643)
AUTHORS Lahn,M., Wade,D., Gram,J.F. and Garvin,M.
TITLE Atp binding cassette transporter protein abc1 polypeptides
JOURNAL Patent: WO 0078971-A 3 28-DEC-2000;
CV THERAPEUTICS, INC. (US)
FEATURES
Location/Qualifiers
1..1643
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 370 a 413 c 457 g 403 t
ORIGIN
Query Match 100.0%; Score 1532; DB 6; Length 1643;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCCTGCTGGTCCACATGCACTTCCAGGCGCTGCTGGCTCTCTAATGGGTC 60
Db 1 GAATTCCTGCTGGTCCACATGCACTTCCAGGCGCTGCTGGCTCTCTAATGGGTC 60
QY 61 TGTCTGAGTCTTGAAGAACCACTGATGTCAGTACCTGGCTGAGCGCTGGGCTGGAGA 120
Db 61 TGTCTGAGTCTTGAAGAACCACTGATGTCAGTACCTGGCTGAGCGCTGGGCTGGAGA 120
QY 121 TCCGTGTTGACTGTAGCATGAGGGGCTTGTACGCTGAATGCTGTATGCAAGTGGTGGG 180
Db 121 TCCGTGTTGACTGTAGCATGAGGGGCTTGTACGCTGAATGCTGTATGCAAGTGGTGGG 180
QY 181 ACTTCTGGAATATGATGAGCTGGAGGTGGGAAGAGAAGTAGGCTTGGGCGAGCTCTC 240
Db 181 ACTTCTGGAATATGATGAGCTGGAGGTGGGAAGAGAAGTAGGCTTGGGCGAGCTCTC 240
QY 241 ATGCCACCTCTATTCGGCCAAACTCAGCTCAAACTGTGAAGAGTCTAAATGTGAATCTG 300
Db 241 ATGCCACCTCTATTCGGCCAAACTCAGCTCAAACTGTGAAGAGTCTAAATGTGAATCTG 300
QY 301 CCTTCAAGTGGGTACAAAGTATCTTTGTCAAGTAGGAGACCTTGTGGCCCTCCAGCT 360
Db 301 CCTTCAAGTGGGTACAAAGTATCTTTGTCAAGTAGGAGACCTTGTGGCCCTCCAGCT 360
QY 361 GCATTCGAGGCGCTGCTGGCTCTTCAAGGCTCTGCTCTGCTGCTGCTCTGCTCTGCT 420
Db 361 GCATTCGAGGCGCTGCTGGCTCTTCAAGGCTCTGCTCTGCTGCTGCTCTGCTCTGCT 420
QY 421 CCTTCAGGCGAGATCATATTTAGACTCTTACAGCTCTTACAGCTTGTGAGTCTGAGTCT 480
Db 421 CCTTCAGGCGAGATCATATTTAGACTCTTACAGCTCTTACAGCTTGTGAGTCTGAGTCT 480
QY 481 AGGTGACATTTAGTTTGTGCTCTGATGAATGACTTAAATATTTAGACATATGTTGTGTA 540
Db 481 AGGTGACATTTAGTTTGTGCTCTGATGAATGACTTAAATATTTAGACATATGTTGTGTA 540
QY 541 GGCCTGCACTTCTACTCTGCTCTTTTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 600
Db 541 GGCCTGCACTTCTACTCTGCTCTTTTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 600
QY 601 CCCCTACAGCCAAAGGCAACAGATAAGTGGAGGCTGGAGGCTGGCTACATAATTTTAC 660
Db 601 CCCCTACAGCCAAAGGCAACAGATAAGTGGAGGCTGGAGGCTGGCTACATAATTTTAC 660
QY 661 ACAGCTGCAATCTCTGCTGCACTTCAAAAGTATACAAACATAAATACAAAGTCTCTG 720
Db 661 ACAGCTGCAATCTCTGCTGCACTTCAAAAGTATACAAACATAAATACAAAGTCTCTG 720
QY 721 TTTTATCACAGGAGGCTGATCAATATAATGAAATTAAGAGGGGCTGGTCCCATATTTG 780
Db 721 TTTTATCACAGGAGGCTGATCAATATAATGAAATTAAGAGGGGCTGGTCCCATATTTG 780
QY 781 TCTGCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 840
Db 781 TCTGCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 840
QY 841 ATGAAGAGAGCAGTAAAGATGTTCTCTGGGCTCTCTGAGGAGCTGGGAGCTCAGGC 900
Db 841 ATGAAGAGAGCAGTAAAGATGTTCTCTGGGCTCTCTGAGGAGCTGGGAGCTCAGGC 900
QY 901 TGGGAATCTCCAGGCACTAGTCCCTATCAAAATCAAAATCAAAATCAAAATCAAAATCA 960
Db 901 TGGGAATCTCCAGGCACTAGTCCCTATCAAAATCAAAATCAAAATCAAAATCAAAATCA 960
```

Db 1 GAATTCCTGCTGGTGGCTCCACATGCACTTCCTCAGGGCTGCTTGGCTCTCTATGGTC 60
 QY 61 TGCCTTCAGTGTGATAGAACACTCACTGAGTACTCTGGCTTGAAGCTGGCTGGAGA 120
 Db 61 TGCCTTCAGTGTGATAGAACACTCACTGAGTACTCTGGCTTGAAGCTGGCTGGAGA 120
 QY 121 TCTGTGTGACTGTAGCATGGAGGGCTGCTGTCAGCTGAATGTCTGTATGTCAGGCTGGG 180
 Db 121 TCTGTGTGACTGTAGCATGGAGGGCTGCTGTCAGCTGAATGTCTGTATGTCAGGCTGGG 180
 QY 181 AGTTCTGGAAATATGATGAGCTGGAGGTGGAGAGAAAGTATGAGCTTGGGCACTCTC 240
 Db 181 AGTTCTGGAAATATGATGAGCTGGAGGTGGAGAGAAAGTATGAGCTTGGGCACTCTC 240
 QY 241 ATGCCACCTCAATCTGGCAAAACTCAGGTCARACTGTGAAGAGTCAATGTGAATCTG 300
 Db 241 ATGCCACCTCAATCTGGCAAAACTCAGGTCARACTGTGAAGAGTCAATGTGAATCTG 300
 QY 301 CCCTTCAAGGTGCTACAAAGGTATCTTTGTCAAGGTAGGAGACCTTGTGGCTCCACGT 360
 Db 301 CCCTTCAAGGTGCTACAAAGGTATCTTTGTCAAGGTAGGAGACCTTGTGGCTCCACGT 360
 QY 361 GCATTCAGGGCTGCTGCTGGCTCTCTACGGGCTGTGCTGAGTCTGCTGAATCTC 420
 Db 361 GCATTCAGGGCTGCTGCTGGCTCTCTACGGGCTGTGCTGAGTCTGCTGAATCTC 420
 QY 421 CCTTCAGGCGAGATTTCATATTAGACTCTTCACAGTTTGACCTGAGTTTGGCCAGAA 480
 Db 421 CCTTCAGGCGAGATTTCATATTAGACTCTTCACAGTTTGACCTGAGTTTGGCCAGAA 480
 QY 481 AGGTGACATTTAGTTGGTGTGATGAACTGAACTAAATATTAGACATATGCTGTGA 540
 Db 481 AGGTGACATTTAGTTGGTGTGATGAACTGAACTAAATATTAGACATATGCTGTGA 540
 QY 541 GGCCTGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 Db 541 GGCCTGCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
 QY 601 CCCCCTACAGCAAGGCAAGAGATAGTTGAGCTCTGAGTCTGGAGTGTACATAATTTAC 660
 Db 601 CCCCCTACAGCAAGGCAAGAGATAGTTGAGCTCTGAGTGTGGAGTGTACATAATTTAC 660
 QY 661 AGGACTGCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
 Db 661 AGGACTGCAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
 QY 721 TTTTATACAGGAGGCTGATCAATATATGAAATTAAGAGGGGCTGCTCCATATTC 780
 Db 721 TTTTATACAGGAGGCTGATCAATATATGAAATTAAGAGGGGCTGCTCCATATTC 780
 QY 781 TTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 840
 Db 781 TTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 840
 QY 841 ATGAGAGAGAGAGTAAATGTTCTCTCGGCTCTCTGAGGAGACCTGGGAGCTCAGGC 900
 Db 841 ATGAGAGAGAGAGTAAATGTTCTCTCGGCTCTCTGAGGAGACCTGGGAGCTCAGGC 900
 QY 901 TGGGAATCTCCAGGAGGAGTGTGCTATCAAAATCAAGTCCAGGTTGTGGGGGA 960
 Db 901 TGGGAATCTCCAGGAGGAGTGTGCTATCAAAATCAAGTCCAGGTTGTGGGGGA 960
 QY 961 AAACAAAGCAGCCCATACCCAGAGGAGTGTCCGCTTCCCTCCCTCCAGCTAGGCC 1020
 Db 961 AAACAAAGCAGCCCATACCCAGAGGAGTGTCCGCTTCCCTCCCTCCAGCTAGGCC 1020
 QY 1021 TTTGAAGGAACAAAGACAGACAAATATGATGGCTCTGAGGAGATTCAGCTAG 1080
 Db 1021 TTTGAAGGAACAAAGACAGACAAATATGATGGCTCTGAGGAGATTCAGCTAG 1080
 QY 1081 AGCTCTCTCCCTCAATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1140
 Db 1081 AGCTCTCTCCCTCAATCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1140

QY 1141 GCGGAAGCAGGATTTAGAGGAAGCAAAATTCACACTGGTGCCTTGGCTGCCGGAAGCTG 1200
 Db 1141 GCGGAAGCAGGATTTAGAGGAAGCAAAATTCACACTGGTGCCTTGGCTGCCGGAAGCTG 1200
 QY 1201 GACTAGAGATCTGCGGCGCAGCCGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 1260
 Db 1201 GACTAGAGATCTGCGGCGCAGCCGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 1260
 QY 1261 GGCCTGCGGCGGGAAGGAGACACAGCCGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 1320
 Db 1261 GGCCTGCGGCGGGAAGGAGACACAGCCGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 1320
 QY 1321 CCCCACCCACCCACCTCCGCCCACTCCCTAGATGTGTGCTGGCGGCTGGAAGCTGCG 1380
 Db 1321 CCCCACCCACCCACCTCCGCCCACTCCCTAGATGTGTGCTGGCGGCTGGAAGCTGCG 1380
 QY 1381 CCGCTTAAAGGG 1440
 Db 1381 CCGCTTAAAGGG 1440
 QY 1441 CAGAGCGCGGGAAGCGGG 1500
 Db 1441 CAGAGCGCGGGAAGCGGG 1500
 QY 1501 TGGGCTCGGTGCGAGCCGATCTATAAAGGAA 1532
 Db 1501 TGGGCTCGGTGCGAGCCGATCTATAAAGGAA 1532

RESULT 3
 AL359182/c
 LOCUS
 DEFINITION Human DNA sequence from clone RP11-217B7 on chromosome 9, complete sequence.
 ACCESSION AL359182
 VERSION AL359182.20 GI:18151453
 KEYWORDS HTG
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Skuce, C.
 Direct Submission
 Submitted (11-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
 On Jan 15, 2002 this sequence version replaced gi:18121468.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at
 http://www.sanger.ac.uk/projects/C.elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr9
 RP11-217B7 is from the library RPO1-11.1 constructed by the group of Pieter de Jong. For further details see

DEFINITION Sequence 1 from Patent WO0115676.
ACCESSION AX092589
VERSION AX092589.1 GI:13444647
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 183999)
Hayden, M.R., Brooks-Wilson, A.R., Pinstone, S.N. and Clee, S.M.
Compositions and methods for modulating hdl cholesterol and
triglyceride levels
Patent: WO 0115676-A 1 08-MAR-2001;
University of British Columbia (CA); Xecon Genetics Inc. (CA)
JOURNAL Location/Qualifiers
FEATURES
Source 1..183999
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 49549 a 37944 c 41170 g 54950 t 386 others
ORIGIN
Query Match 90.6%; Score 1387.6; DB 6; Length 183999;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 1503; Conservative 2; Mismatches 10; Indels 23; Gaps 9;
QY 16 GCTCCAGATCACTTCAGGSCCTGCTGG--CTCTCTATGGTCTGCTCGAGTGT 73
Db 27234 GCTCCAGATCACTTCAGGSCCTGCTGGCCCTCTATGGTCTGCTCGAGTGT 72933
QY 74 GATGAACACTGATGTGAGTACCTGGCTTGAG--CGTGGCTGGAGATCCCTGTTGACTG 132
Db 27294 GATGAACACTGATGTGAGTACCTGGCTTGAGCGCTGGCTGGAGATCCCTGTTGACTG 27353
QY 133 TAGCATGAGGGGGCTTGT-CAGCTGAATGCTGTATGAGTGTGCTGGAGTCTGGAC 191
Db 27354 TAGCATGAGGGGGCTTGTGAGCTGAATGCTGTATGAGTGTGCTGGAGTCTGGAC 27413
QY 192 ATGATGGAGCTGGAGTGGGAGAGAAGTAGGCTTGGGCGAGCTCTCATGCCACCTCA 251
Db 27414 ATGATGGAGCTGGAGTGGGAGAGAAGTAGGCTTGGGCGAGCTCTCATGCCACCTCA 27473
QY 252 TTCGGCCAAACTCAGTCAAACTGTGAAGTCTAAATGTAATCTGCTTCAAGT 311
Db 27474 TTCGGCCAAACTCAGTCAAACTGTGAAGTCTAAATGTAATCTGCTTCAAGT 27533
QY 312 GGTCAAAAGGTATCTTGTCAAGTAGGAGACCTTGTGGCTCCACGTGCACTTCCAGG 371
Db 27534 GGTCAAAAGGTATCTTGTCAAGTAGGAGACCTTGTGGCTCCACGTGCACTTCCAGG 27593
QY 372 GCTGCTT-GGCTCTTAGGGCTGTGCTGAGTCTTCTAIGAA--TCTCCCTTCAGG 428
Db 27594 GCTGCTTGGGCTCTTCTACGGTCTGCTGAGTCTTCTAIGAACTGCTCTCAGG 27653
QY 429 GCAGATCATATTAGACTCTACAGTTCACCTGAGTTGACCTGAGTTTGGCCAGAAATAGGTGACA 488
Db 27654 GCAGATCATATTAGACTCTACAGTTCACCTGAGTTGACCTGAGTTTGGCCAGAAATAGGTGACA 27713
QY 489 TTAGTCTGCTGCTGATGAATGACTAAATATTAGACATATGTTGTAGGCTGCA 548
Db 27714 TTAGTCTGCTGCTGATGAATGACTAAATATTAGACATATGTTGTAGGCTGCA 27773
QY 549 TTCCTACTCTGCTCTTTTTCCTCCCTCCAGTGTGTTGGGTAGTTTGTGCTCCCTTAC 608
Db 27774 TTCCTACTCTGCTCTTTTTCCTCCCTCCAGTGTGTTGGGTAGTTTGTGCT-CCCTTAC 27832
QY 609 AGCCAAAGSCAAACAGATAAGTTGAGTCTGGAGTCTGCTAGATATTTCACCACTGC 668
Db 27833 AGCCAAAGSCAAACAGATAAGTTGAGTCTGGAGTCTGCTAGATATTTCACCACTGC 27892
QY 669 AATTCCTGCTGCTGCTACCAAAATGTATACAAATGTATACAAATGTATACAAATGTATAC 728
Db 27893 AATTCCTGCTGCTGCTACCAAAATGTATACAAATGTATACAAATGTATAC 27952

QY 729 ACAGGAGGCTGATCATATATATGAATAAAGAGGGGCGTGTCCATATGTTCTGTCTGT 788
Db 27953 ACAGGAGGCTGATCATATATATGAATAAAGAGGGGCTGTGT-CCATATGTTCTGTCTGT 28011
QY 789 TTTTGTGTTGTTGTTCTTTT-----TTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 839
Db 28012 TTTTGTGTTGTTGTTCTTTNN 28071
QY 840 TATGAGAGAGAGCAGTAAAGATGTTCTCTCGGGTCTCTGAGGAGCCTGGGAGCTCAGG 899
Db 28072 TATGAGAGAGAGCAGTAAAGATGTTCTCTCGGGTCTCTGAGGAGCCTGGGAGCTCAGG 28131
QY 900 CTGGGAATCTCCAGGAGCTAGTCTGCTATCAAAATCAAAAGTCCAGTGTGTTGGGGGG 959
Db 28132 CTGGGAATCTCCAGGAGCTAGTCTGCTATCAAAATCAAAAGTCCAGTGTGTTGGGGGG 28191
QY 960 AAACAAAGAGAGCCTATTAATCCAGAGAGTCTCCGCTCTCCCTACCCAGCCTAGGC 1019
Db 28192 AAACAAAGAGAGCCTATTAATCCAGAGAGTCTCCGCTCTCCCTACCCAGCCTAGGC 28251
QY 1020 CTTTAAAG 1079
Db 28252 CTTTAAAG 28311
QY 1080 GAGCTCTCTCTCCCTCAATCCCTCCCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1139
Db 28312 GAGCTCTCTCTCCCTCAATCCCTCCCTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 28371
QY 1140 TCCGAAAG 1199
Db 28372 TCCGAAAG 28431
QY 1200 GAGCTAGAGAGTCTCCGAG 1259
Db 28432 GAGCTAGAGAGTCTCCGAG 28491
QY 1260 GGGCCGGGGGGGAG 1319
Db 28492 GGGCCGGGGGGGAG 28551
QY 1320 A-----CCCCCACCACCCACCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1374
Db 28552 ANNNNNCCCCCACCACCCACCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 28611
QY 1375 CTTCCCTGTTTAAAGGG 1434
Db 28612 CTTCCCTGTTTAAAGGG 28671
QY 1435 CATAACAG 1494
Db 28672 CATAACAG 28731
QY 1495 AACCTCTGGCTGCTGAG 1532
Db 28732 AACCTCTGGCTGCTGAG 28769

RESULT 5
AX351031
LOCUS AX351031
DEFINITION Sequence 3 from Patent WO0183745.
ACCESSION AX351031
VERSION AX351031.1 GI:18616387
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
Rosier-Montus, M.F., Prades, C., Lemoine, C., Naudin, L., Benefle, P.,
Brewer, B., Duvergier, N., Remaley, A. and Santamarina-Fojo, S.
Regulatory nucleic acid sequences of the abcl gene
Patent: WO 0183745-A 3 08-NOV-2001;
JOURNAL

linear

DNA

2993 bp

PAT 06-FEB-2002

```
FEATURES             Aventis Pharma S.A. (PR)
Source               Location/Qualifiers
1..2893
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT          745 a 681 c 769 g 696 t
ORIGIN
Query Match          89.7%; Score 1374.4; DB 6; Length 2893;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 1504; Conservative 0; Mismatches 6; Indels 26; Gaps 10;
QY 16 GCTCCACATGACATTCAGGGCCGCTGG--CCCTTCATGGGCTGCTGCTGAGTGT 73
DB 1345 GCCTCCACATGACATTCAGGGCCGCTGGCTTCTATGGCTCTGCTGCTGAGTGT 1404
QY 74 CATAGAACCACTGATGAGTACCTGGGCTTGAG--CGTGGGCTGGAGATCCCTGTTGACG 132
DB 1405 GATAGAACCACTGATGAGTACCTGGGCTTGAGCCGCTGAGCCGCTGCTGTTCACTG 1464
QY 133 TAGCATGGAGGGGCTGCT--CAGCTGAATGCTGATGACAGTGGTGGAGTCTGGAAT 181
DB 1465 TAGCAGYAGAGGGGCTGTTGGAGCTGAATGCTGATCAGAGTGGTGGAGTCTGGAAT 1524
QY 192 ATGATGGAGCTGGAGTGGGGAAGAGATGAGCTTGGGCGACCTCTCTCATGCCACCTCA 251
DB 1525 ATGATGGAGCTGGAGTGGGGAAGAGATGAGCTTGGGCGACCTCTCTCATGCCACCTCA 1584
QY 252 TTCTGGCCAAACACTCAGTCAAACTGTGAAGAGTCTAAATGTGAATCTGCCCTCAAGT 311
DB 1585 TTCTGGCCAAACACTCAGTCAAACTGTGAAGAGTCTAAATGTGAATCTGCCCTCAAGT 1644
QY 312 GCTCAAAAGGTAATCTTGTCAAGGTAGAGACCTGTG--GGCCCTCCAGCTGCACTTCCAGG 371
DB 1645 GCCTCAAAAGGTAATCTTGTCAAGGTAGAGACCTGTG--GGCCCTCCAGCTGCACTTCCAGG 1704
QY 372 GCCTCCTT--GGCCTCTTACGGGCTCTGCTCGATGCTCTTATGAATCTCCCTTCAGGGC 430
DB 1705 GCCTCCTTGGGCTCTTACGGGCTCTGCTCGATGCTCTTATGAATCTCCCTTCAGGGC 1761
QY 431 AGATTCATATTTAGACTCTTCACAGTGTGACCTGAGTTTGGCCAGATAAGTGCAGATT 490
DB 1762 AGATTCATATTTAGACTCTTCACAGTGTGACCTGAGTTTGGCCAGATAAGTGCAGATT 1821
QY 491 TAGTTTGTGGCTTGATGAATGACHTAAATATTTAGACATATGGTGTGAGSCCTCCATT 550
DB 1822 TAGTTTGTGGCTTGATGAATGACHTAAATATTTAGAC--ATGGTGTGAGSCCTCCATT 1879
QY 551 CCTACTCTTGCTTTTGTGCCCCCTCCAGTGTGTTGGGTAGTGTGCTCCCTCCCTACAG 610
DB 1880 CCTACTCTTGCTTTTGTGCCCCCTCCAGTGTGTTGGGTAGTGTGCT--CCCTCAGAG 1938
QY 611 CCAAGGCAACACAGATAAGTTGGAGTCTGGAGTGGCTGACATAAATTTACAGCACTGCA 670
DB 1939 CCAAGGCAACACAGATAAGTTGGAGTCTGGAGTGGCTGACATAAATTTACAGCACTGCA 1998
QY 671 TTCTGTGGCTGCACHTCACAAATGATACAACTAAATACAAAGTCTGTTTATCAC 730
DB 1995 TTCTGTGGCTGCACHTCACAAATGATACAACTAAATACAAAGTCTGTTTATCAC 2058
QY 731 AGGGAGGCTGATCAATATAATCAATTAAGAGGGGCTGCTCCCATATGTTCTGTGTT 790
DB 2059 AGGGAGGCTGATCAATATAATCAATTAAGAGGGGCTGCT--CCATATGTTCTGTGTT 2117
QY 791 TTC-----TTTGTGTTGTTCTTTTGTGTTTGTGGGCTCTCTCTCTCAATTTA 841
DB 2118 TTGTTGTGTTGTTGTTGTTCTTTTGTGTTTGTGGGCTCTCTCTCTCAATTTA 2177
QY 842 TGAAGAGAGCAGTAAAGTTCCTCTCTCGGTCCTCCGTCGAGGACCTGGGAGCTCAGGCT 901
DB 2178 TGAAGAGAGCAGTAAAGTTCCTCTCTCGGTCCTCCGTCGAGGACCTGGGAGCTCAGGCT 2237
QY 902 GGGATCTCCAGGCGAGTGGTGGCTATCAAAATCAAAAGTCCAGGTTTGTGGGGGAA 961
```

```
DB 2238 GGGAAATCTCCAAGGAGTAGGTGCGCTATCAAAAATCAAGTCCAGGTTTGTGGGGGAA 2297
QY 962 AACAAAGCAGCCCATTTACCCAGAGAGTGTCCGGCTTCCCGCTCACCCAGCTAGGCT 1021
DB 2298 AACAAAGCAGCCCATTTACCCAGAGAGTGTCCGGCTTCCCGCTCACCCAGCTAGGCT 2357
QY 1022 TTGAAGGAACAAAGACACAGACAAATGATTGGCTGCTGAGGAGATTTCAGCCTAGA 1081
DB 2358 TTGAAGGAACAAAGACACAGACAAATGATTGGCTGCTGAGGAGATTTCAGCCTAGA 2417
QY 1082 GCTCTCTCTCCCGCAATCCCTCCCGCTGAGCAAACTACAAAGGAAAAAAATTTG 1141
DB 2418 GCTCTCTCTCCCGCAATCCCTCCCGCTGAGCAAACTACAAAGGAAAAAAATTTG 2477
QY 1142 CGAAAGCAGGATTAGAGGAAGCAAAATTCACACTGGTGGCTCCCGGGGACG--GG 1201
DB 2478 CGAAAGCAGGATTAGAGGAAGCAAAATTCACACTGGTGGCTCCCGGGGACG--GG 2537
QY 1202 ACTAGAGAGTCTGCGCGCAGCCCGAGCCGAGCCCTTCCCGCGGCTCTTAGCGCGCGG 1261
DB 2538 ACTAGAGAGTCTGCGCGCAGCCCGAGCCGAGCCCTTCCCGCGGCTCTTAGCGCGCGG 2597
QY 1262 GCCCGGGCGGGGGAAGGAGGAGCAGACCGCGCACCCCTAAGACACCTGTGTACCTCCAC 1321
DB 2598 GCCCGGGCGGGGGAAGGAGGAGCAGACCGCGCACCCCTAAGACACCTGTGTACCTCCAC 2657
QY 1322 -----CCCCACCCACCCACCTCCCGCACTCCCGCACTTCCCGCGGCTCTTAGCGCGCGG 1376
DB 2658 CCCCACCCACCCACCCACCTCCCGCACTTCCCGCACTTCCCGCGGCTCTTAGCGCGCGG 2717
QY 1377 TCGCCCGTTTAAAGGGCGGGCCCGCTCCACGTCCTTCTGCTGAGTGAAGTACTACA 1436
DB 2718 TCGCCCGTTTAAAGGGCGGGCCCGCTCCACGTCCTTCTGCTGAGTGAAGTACTACA 2777
QY 1437 TAAACAGAGCGCGGGAACCGGGCGGGGAGGAGGAGACACAGCTTTTACCGATAGTAA 1496
DB 2778 TAAACAGAGCGCGGGAACCGGGCGGGGAGGAGGAGACACAGCTTTTACCGATAGTAA 2837
QY 1497 COTCTGCGCTCGTCCAGCGCAATCTATAAAGGA 1532
DB 2838 COTCTGCGCTCGTCCAGCGCAATCTATAAAGGA 2873
RESULT 6
AX351029 LOCUS AX351029 3231 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 1 from Patent WO0183746.
ACCESSION AX351029
VERSION AX351029.1 GI:18618385
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1.
AUTHORS Rosier-Montus,M.F., Prades,C., Lemoine,C., Naudin,L., Denefle,P.,
Brewer,B., Duverger,N., Remaley,A. and Santamarina-Pofo,S.
TITLE Regulatory nucleic acid sequences of the abcl gene
JOURNAL Patent: WO 0:83746-A 1 08-NOV-2001;
FEATURES
Location/Qualifiers
1..3231
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 809 a 773 c 876 g 773 t
ORIGIN
Query Match 89.7%; Score 1374.4; DB 6; Length 3231;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 1504; Conservative 0; Mismatches 6; Indels 26; Gaps 10;
QY 16 GGCTCCACATGACATTCAGGGCCGCTGG--CCTTCTATGGCTGCTGCTGAGTGT 73
```



```

110646..110866,112124..112196,113183..113385,
115053..115101,115297..115470,116806..116954,
118604..118728,119864..120062,123094..123283,
124804..124898,126208..126240,127363..127468,
128943..129017,129539..129708,130939..131116,
133114..133229,133341..133485,134526..134649,
135736..135865,136130..136250,137038..137100,
138008..138114,140469..140610,140982..141116,
142060..142163,142646..142738,143397..143640,
144381..1448034)
/gene="ABCA1"
/product="ABCA1"
join(1454..1674,25831..25923)
/gene="ABCA1"
2406..2678
/rpt_family="Alu"
repeat_region
4240..4509
/rpt_family="Alu"
repeat_region
4998..5270
/rpt_family="Alu"
repeat_region
5563..5839
/rpt_family="Alu"
repeat_region
6615..6877
/rpt_family="Alu"
repeat_region
8800..9046
/rpt_family="Alu"
repeat_region
9998..10279
/rpt_family="Alu"
repeat_region
11859..12133
/rpt_family="Alu"
repeat_region
12810..12902
/rpt_family="Alu"
repeat_region
15220..15399
/rpt_family="Alu"
repeat_region
15508..15787
/rpt_family="Alu"
repeat_region
18600..18860
/rpt_family="Alu"
repeat_region
20950..21206
/rpt_family="Alu"
repeat_region
25038..25314
/rpt_family="Alu"
repeat_region
CDS
join(25924..25989,40385..40478,45012..45153,46423..46541,
67729..67850,70831..71007,83960..84052,89010..89250,
91962..92101,92433..92549,96758..96955,97702..97907,
98428..98604,100391..100613,102360..102581,103642..103846,
104951..105064,106862..107033,108023..108154,
109460..109602,109806..109943,110646..110866,
112124..112196,113183..113385,115053..115101,
115297..115410,116806..116954,118604..118728,
119964..120062,123094..123283,124804..124898,
126208..126240,127363..127468,128943..129077,
129539..129708,130939..131116,133114..133229,
133341..133485,134526..134649,135736..135865,
136130..136250,137038..137100,138008..138114,
140469..140610,140982..141116,142060..142163,
142646..142738,143397..143640,144381..144721)
/gene="ABCA1"
/codon_start=1
/product="ABCA1"
/protein_id="AAF65276.1"
/db_xref="GI:9247086"
/translation="MACPOLRLLLWKKLRRRTQCOLLEWAWPLFPLILISVRL
SYPEQHECHFPKAMPAGILPWQIICNANPCFRYPPTCEAPGVGNKNSV
ARLSDARRLLNYSOKTSMKDMKRVLETQQIKSSSNLKLOFLVDNETFSGFLYH
NLSLPKTVDKMLRADVLEKVFLOGQLHLTSCNGSKSEMIQLQDQVESELCGLP
REKLAARVLRNSMDILKPIRLTNSTSPFSPKELAZATKLLHSLGTAEQLFSMR
SWSDQREVFLLTNVSSSSITQIQAIVSRVCGHPGCGLKIKSLNWEYDNKYKALF
GGNCTEDAEATFYDNSTTPYCNLDLKNLESSPLSIWKALKPLLVGKIATYDPIPAT
ROYMAEVNKTFOELAVFLDGLGMFEELSPKIWTFMENSCNDLYMLLDSDHFWNE
QQLDGLDWTADIVAFIAKAFEDFVQSSNGSVYTWREAFNQTQAIRTISRMECVNLN
KLELATVEWLKNSMELLDERKEWAGIVFTGILFGSIELPHVYKIRMDIDNVERT
NKIKDGYWDPGPRDPEDMEYVWGGFAYLODVVEQAIIRVLGTERTKKTGYVMQMPY

```

```

PCYVDDIELRVMSBMLEMTLAWIVYAVIILKIGIYVEKEARLKETMRIMGLDNLILW
PSWFLSILPLLVSAGLIVLILKGLNLLPSPDPSVFFVSFAVAVTILQCFLLSTILF
SRANLAACGGIIFTLYLPIVLCVAMCDYVFTLKIIFAXLLSPVAFGFGCEYFALFE
EOGIGVQMDNLFPSPVEEDGNLTYSMMLEFDTLXGVTWYIYAVPGGYGIPRW
YFCTKSYWFGESDEKSHPSNKRISCEIMBEPEHLKLGVSIGNLVKYYRGMKV
AVDGLALNVEGQITSLFGNGAGKTTMSLTGLTLPPTSTAVILQKDIRSEMTIR
ONLACPOHNVLFDMIAVESHIFWYASLKGISEKHVAEMOMALDYGLEPSSKLSKT
SOLSGGMORLSVALAFVGSKVILDEPTAGVPPYSGRIWELLILAYROGRTIILST
HMDEADVLDRIAIISHGKCCVGSFLFKNQIGTGYITLVKKNVSSLSLSCNRS
STYSYLKEDSVSQSSDAGLGSJESDITLIDVSATSLNLRKHVSARLVEDIGHEL
TYLVPYAAKEGAEVLFHEIDRLSDIGSSYSISPTLEIEIFLVAEBSGVDAE"TS
DGLTPARRNRRAFGDKOSCLRPPTEDDAADPNDSDIDPESRETLLSGMDGKGYOVK
GWKTLOQFVALLWKRLLIAPRSKGFQAQIVLPAVGVIALVPSLIVPFGKPSLE
LOPMYNEQYTFVNSDAPEDGTLELLNALTKDPGCTGTRCHGNPDPDTCQAGEEW
TTAPVQITKIDFONGNWTQNPSPACSSDKIKMPLPYCPGAGLPPFPQKQNTA
DILQDLGRNISDYLVKTYVOITAKSLKNKIWNENFRYGGFSLGVSNTQALPPSEYN
DAXKMKHLKLAKDSSADRLNSLGRFMTGLDTRNNVKYFNKNKGWHAISSFLNIN
NAILRANLQGENSHYGIATFNHPLNLTQOLSEVAXMTTSDVLVSIQVIFAMSV
PASTVFLIQUERSVAKHLOFISGVKPVIVLWSNFVMDMNVVVPAT-VIIIFCFQO
KSYVSTNLPLVALILLLYGWSITPLMPASTVFKIPSTAYVLTSLVNLFIGINGSVA
TFVLELFDKLNINIDILKSVLIFPHFCGLGRGLIDMVKNQAMADALEFGENRFS
RLSWDLVGRNLFAMAVEGVFFLTIVLIQYRFFIRPRVNAKLSPLENDEDEDVREHO
RILDGGQNDILEIKELTKIYRRKRKPAVIRICVGPCEGGLLYNGAGKSTIFKH
LTGDTTVTRGDALFNXSLISNIHEVQNMGYCPQFDATTELLTGREHVERFALLRGV
PEKVGKVGWAIKRLGLVYKERYAGNYSGNKRLSTAMALIGGPPVFLDPTTG
MDPKARFLNVALSVYKGPSVLTSSMBECCALATRKALWNGRCLGSGOHLK
NRFQDGYTVIRIAGSPDLKPQVDFGLAPFGVYXKEKHNMLQYQLPSSLSSLAK
PSTLSQSKKRHLIEDYSVPSQTLQOVFNFAKQSDDDHLDLSLHKNOIVDVAVLT
SFLQDEKVKESYV"
29916..30180
/rpt_family="Alu"
31050..31314
/rpt_family="Alu"
31692..31973
/rpt_family="Alu"
36167..36440
/rpt_family="Alu"
37617..37892
/rpt_family="Alu"
37999..38166
/rpt_family="Alu"
38172..38269
/rpt_family="Alu"
38995..39271
/rpt_family="Alu"
41547..41565
/gene="ABCA1"
41593..41835
/note="A009X28"
41657..41679
/rpt_family="Alu"
/gene="ABCA1"
/note="A009X28"
42320..42584
/rpt_family="Alu"
44349..44526
/rpt_family="Alu"
48473..48669
/rpt_family="Alu"
50547..50614
/rpt_family="Alu"
54285..54418
/rpt_family="Alu"
54849..54988
/rpt_family="Alu"
55029..55226
/rpt_family="Alu"
55297..55344
/rpt_family="Alu"
58501..58935
/note="LIR"
/rpt_family="HERV"
51246..61489

```

Db	777	GGGACCTGGGGAGCTCAGGCTGGGAATCTCCAGAGCAGTAGTGCGCTATCAAAATCAA	836
QY	941	AGTCCAGGTTGTGCGGGGAAAACAAAAGCAGGCCATTACCCAGAGGACTGFCGCCCTTC	1000
Db	837	AGTCCAGGTTGTGCGGGGAAAACAAAAGCAGGCCATTACCCAGAGGACTGFCGCCCTTC	896
QY	1001	CCCTCACCCAGCCTAGSCCTTTCAGAGGAAACAAAAGACACAGACAAATGATTCGGCTTC	1060
Db	897	CCCTCACCCAGCCTAGSCCTTTCAGAGGAAACAAAAGACACAGACAAATGATTCGGCTTC	956
QY	1061	CTGAGGAGAGATTACGCGCTAGAGCTCTCTCTCCGCCAA1CCCTCCCTCCGGGTGAGGAAC	1120
Db	957	CTGAGGAGAGATTACGCGCTAGAGCTCTCTCTCCGCCAA1CCCTCCCTCCGGGTGAGGAAC	1016
QY	1121	TAAACAAGGAAAAAATAATTCGGAAACACAGGATTATAGAGGACAAATTCACHTGTGC	1180
Db	1017	TAAACAAGGAAAAAATAATTCGGAAACACAGGATTATAGAGGACAAATTCACHTGTGC	1076
QY	1181	CCCTGGCTGCCGGGAAACGTGTGACATAGAGAGTCTCGCGGCGAGCCCGCAGCCAGCGCTC	1240
Db	1077	CCCTGGCTGCCGGGAAACGTGTGACATAGAGAGTCTCGCGGCGAGCCCGCAGCCAGCGCTC	1136
QY	1241	CCGGCGGCTCTTAGCGCGCGCGGCCCGGGGGAAGGAGCGACAGCCGGGACCCCTAA	1300
Db	1137	CCGGCGGCTCTTAGCGCGCGCGGCCCGGGGGAAGGAGCGACAGCCGGGACCCCTAA	1196
QY	1301	GACACCTGCTCTACCTCCAC-----CCCCACCCACCACCTCCGCCCACTCCCTAGA	1355
Db	1197	GACACCTGCTCTACCTCCAC-----CCCCACCCACCACCTCCGCCCACTCCCTAGA	1256
QY	1356	TGTCGTGCGGCGCTGAACTGCGCCCTTTAAAGGGGGGGCGCCGGGTCCACGTGCTTT	1415
Db	1257	TGTCGTGCGGCGCTGAACTGCGCCCTTTAAAGGGGGGGCGCCGGGTCCACGTGCTTT	1316
QY	1416	CTGCTGAGTCTGACTGAACTACATACACAGCCCGGGAACGGGGGCGGAGGAGGAGAC	1475
Db	1317	CTGCTGAGTCTGACTGAACTACATACACAGCCCGGGAACGGGGGCGGAGGAGGAGAC	1376
QY	1476	ACAGGCTTTGACCGGATAGTAACTCTGGGCTGGGTGACGCCCAATCTATAAAGGAA	1532
Db	1377	ACAGGCTTTGACCGGATAGTAACTCTGGGCTGGGTGACGCCCAATCTATAAAGGAA	1433
RESULT	8		
LOCUS	AF287262	201344 bp DNA linear	PAI 29-APR-2003
DEFINITION	Homo sapiens ATP-binding cassette 1 sub-family A member 1 (ABCA1) and SNAP protein genes, complete cds.		
ACCESSION	AF287262		
VERSION	AF287262.1	GI:13876612	
KEYWORDS	Homo sapiens.		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 201144)		
TITLE	Qiu,Y., Caveller,L., Chiu,S., Yang,X., Rubin,E. and Cheng,J.F. Human and mouse ABCA1 comparative sequencing and transgenesis studies revealing novel regulatory sequences		
JOURNAL	Genomics 73 (1), 66-76 (2001)		
MEDLINE	21251004		
PUBMED	11352567		
REFERENCE	2 (bases 1 to 201144)		
AUTHORS	Qiu,Y., Caveller,L., Chiu,S., Rubin,E. and Cheng,J.F.		
TITLE	Direct Submission		
JOURNAL	submitted (13-JUL-2000) Genome Science Department, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, MS 84-171, Berkeley, CA 94720, USA		
FEATURES	Location/Qualifiers		
source	1..201144		
gene	/organism="Homo sapiens" /db_xref="taxon:9606" 33931..181457		

```

mRNA
join(33931..34151,58319..58476,72893..72986,77520..77661,
78931..79049,101048..101169,104328..104328,117287..117379,
123239..123269,125285..125424,125754..125870,
130079..130276,131023..131228,131749..131925,
133708..133930,135677..135898,136959..137163,
138268..138381,140179..140350,141340..141471,
142777..142919,143123..143260,143363..144183,
145441..145513,146657..146859,148527..148575,
148771..148884,150280..150428,152078..152202,
153438..153536,156568..156757,158278..158372,
159682..159714,160837..160942,162417..162491,
163013..163182,164413..164590,166588..166703,
168815..168959,168505..168828,169715..169844,
170109..170229,171017..171079,171987..172093,
174448..174589,174961..175095,176039..176142,
176625..176717,177376..177619,178560..181457)
/genes="ABCA1"
/product="ATP-binding cassette 1 sub-family A member 1"
33931..34151
/genes="ABCA1"
/number=1
58319..58476
/genes="ABCA1"
/number=2
join(58411..58476,72893..72986,77520..77661,78931..79049,
101048..101169,104328..104328,117287..117379,
123239..123269,125285..125424,125754..125870,
130079..130276,131023..131228,131749..131925,
133708..133930,135677..135898,136959..137163,
138268..138381,140179..140350,141340..141471,
142777..142919,143123..143260,143363..144183,
145441..145513,146657..146859,148527..148575,
148771..148884,150280..150428,152078..152202,
153438..153536,156568..156757,158278..158372,
159682..159714,160837..160942,162417..162491,
163013..163182,164413..164590,166588..166703,
168815..168959,168505..168828,169715..169844,
170109..170229,171017..171079,171987..172093,
174448..174589,174961..175095,176039..176142,
176625..176717,177376..177619,178560..181457)
/genes="ABCA1"
/codon_start=1
/product="ATP-binding cassette 1 sub-family A member 1"
/protein_id="AAK43526.1"
/db_xref="GI:13876613"
/translation="MACPQLKLLWNLTFRRTQTCOLLLEAVPLIFILLSVRL
SYPEQEHCFFPKAMPAGTLEWVGIIICNANPCFRPTGGEAFVGVGNFKSIV
ARLTSARRLLLYSOKDTSMKDMKVLRILOQIKSSNLKIQDFLVNDFSGFLYH
NLSLPKSTVDKMLRADVILHKVFLQGYQLPLSLCNGSKSEEMQLGDQVSELCGLP
KEKLAARVLRSNWDLKPLIRLTINSTPTPSKELAEATKTLHSLGILAELEFSMR
SWSMRQVEMFLINWSSSSQIYQAVSRVCGHPGGGLKIKSLNWDNNKALF
GGNTEDAETFYDNSTPYCNDLKNLESSLRSRIWKALKP.LLVGKILYTPDIPAT
RQVAENKNTFOELAVFDELEGWEELESPKWTFMENSEMDLVMLLDSRDNDHFE
QQLDGLWADQIVIAFLAKHPEDYQSSNGSVYTWREAETNETNOAIRTISRMEQVNLRT
KLEPATEVWFLINKSMELLDERKFWAGIVFTGIPGSELPHHVKYKIRMDIDAVNT
NKIKDWDPRADPPEDMRYVWGCPALQOVVEQALIRVLTGTEKKTGYVMOMPY
PCYVDIPFLVRSRSMPLMTLWLIYSAVILKIGVIEKARLEIRIMGLDNLILW
FWSYTSLLP.LLVAGLLVILKGNLPIYSDPSVVFSLVFAVVT.LQGLFLSTLF
SRANLAACGGIITFTLLPYLVCVAMODYGFTLKIFASLSPVAFGGEGCEALFE
EQIGVQNDLFPESVEEDGNLITVSVMMLNFTFLYGMWTWYIEAVPPGGYGLPRPW
YEPCTKSYRGESEDEKSHPGSNOKRISEICMEEPETHLKLGVSTONLVKYVRDMKV
AVDGLALNPFEGQITSLFHGAGAKITTSILTGLFLPPTSCYATILGADIRSEMTIR
QNLGCPQENLMTVEETVEEHLIFARLGLSKKHVKMEOMA.DVGLPSSKUSKT
SOLSCGMOKLISVALFVGGSKVWILDEPTAGVDYSRG.WELLKYRQGTIILT
HMDADVLGRIALISHGKLCVGSLLFKNLGTGYLTLVKKDVSSSLSCRNSS
SVSYLVKPFDSVQSSDAGLGSDESDITJDVSAISNLIRKHVSEARLYDIGEEJ
TVLYPEARKKAGAFVELPHEIDRLSDIGSYGISETTLEIFILKVAEESVDAPTS
DGLTPARKNRNFAFGKQSLRPFTEDDAADPNDSIDIPESRETDLLSGMDKGSQYVK
GKLTQOQFVALLKRLIARRSRGFFAQIVLPVAFVCIALVSLIIVPPFGKPSLE
LQPMWNYQYTVFSVSDAPEDTGTLELLNLT KDPGGTCRMEGNIPDITPOAGEEW
TTAPVPQTIMDLFQNGNWTMNPSPACQSSDKIKMLPVCPPGAGGLPPQKQNTA
DILQDITGRNISCLVKTYYQIIRAKSLNKKIWNVEFRYGGFSLGVSNQTQALPPSQEVN

```



```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
- (bases 1 to 175064)
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens, clone RP11-1M10
JOURNAL Unpublished
AUTHORS
2 (bases 1 to 175064)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, M., Anderson, M.,
Balwin, J., Bana, N., Becker, R., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Castle, A., Collins, M., Collins, S., Collamore, A.,
Cooke, P., DeArelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferrara, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galan, J., Gady, S., Grant, S., Hagos, B., Heaford, A., Horton, L.,
Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwen, P., McGurk, A., McKernan, K., McLaughlin, J., McElrim, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessier, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, N.
Direct Submission
Submitted (21-Oct-1999), Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 22, 2000 this sequence version replaced gi:6454033.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2510
Center clone name: 1.M.10
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 117571 bases at least Q40
Consensus quality: 145749 bases at least Q30
Consensus quality: 160940 bases at least Q20
Insert size: 145000; agarose-fp
Insert size: 171264; sum-of-contigs
Quality coverage: 2.9 in Q20 bases; agarose-fp
Quality coverage: 3.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1003: contig of 1003 bp in length
* 1004 1103: gap of 100 bp
* 1104 2634: contig of 1531 bp in length
* 2635 2734: gap of 100 bp
* 2735 4415: contig of 1681 bp in length
* 4416 4515: gap of 100 bp
* 4516 5785: contig of 1270 bp in length
* 5786 5885: gap of 100 bp
* 5886 7879: contig of 1994 bp in length
* 7880 9786: gap of 100 bp
* 9787 9786: contig of 1707 bp in length
* 9787 12253: contig of 2467 bp in length
* 12254 12353: gap of 100 bp
* 12354 15228: contig of 2875 bp in length
* 15229 15328: gap of 100 bp
* 15329 17200: contig of 1872 bp in length

```

```

* 17201 17300: gap of 100 bp
* 17301 20131: contig of 2831 bp in length
* 20132 20231: gap of 100 bp
* 20232 22587: contig of 2356 bp in length
* 22588 22687: gap of 100 bp
* 22688 25707: contig of 3020 bp in length
* 25708 25807: gap of 100 bp
* 25808 28184: contig of 2377 bp in length
* 28185 28284: gap of 100 bp
* 28285 31338: contig of 3054 bp in length
* 31339 31438: gap of 100 bp
* 31439 34299: contig of 2861 bp in length
* 34300 34399: gap of 100 bp
* 34400 36318: contig of 3919 bp in length
* 36319 38418: gap of 100 bp
* 38419 42835: contig of 4417 bp in length
* 42836 42935: gap of 100 bp
* 42936 45448: contig of 2513 bp in length
* 45449 45548: gap of 100 bp
* 45549 48116: contig of 2568 bp in length
* 48117 48216: gap of 100 bp
* 48217 52618: contig of 4402 bp in length
* 52619 52718: gap of 100 bp
* 52719 56592: contig of 3874 bp in length
* 56593 56692: gap of 100 bp
* 56693 59635: contig of 2943 bp in length
* 59636 59735: gap of 100 bp
* 59736 63661: contig of 3926 bp in length
* 63662 63761: gap of 100 bp
* 63762 68437: contig of 4576 bp in length
* 68438 68537: gap of 100 bp
* 68538 71458: contig of 2921 bp in length
* 71459 71558: gap of 100 bp
* 71559 76888: contig of 5330 bp in length
* 76889 76988: gap of 100 bp
* 76989 82113: contig of 5125 bp in length
* 82114 82213: gap of 100 bp
* 82214 88220: contig of 6007 bp in length
* 88221 88320: gap of 100 bp
* 88321 93499: contig of 5179 bp in length
* 93500 93599: gap of 100 bp
* 93600 97901: contig of 4302 bp in length
* 97902 98001: gap of 100 bp
* 98002 103016: contig of 5015 bp in length
* 103017 103116: gap of 100 bp
* 103117 109178: contig of 6062 bp in length
* 109179 109278: gap of 100 bp
* 109279 117307: contig of 8029 bp in length
* 117308 117407: gap of 100 bp
* 117408 124079: contig of 6672 bp in length
* 124080 124179: gap of 100 bp
* 124180 131281: contig of 7102 bp in length
* 131282 131381: gap of 100 bp
* 131382 138059: contig of 6878 bp in length
* 138060 138159: gap of 100 bp
* 138160 145491: contig of 7332 bp in length
* 145492 145591: gap of 100 bp
* 145592 157391: contig of 11800 bp in length
* 157392 157491: gap of 100 bp
* 157492 175064: contig of 17573 bp in length.
FEATURES
Location/Qualifiers
1..175064
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-1M10"
/clone_lib="RFCI-11 Human Male BAC"
misc_feature 1..1003
/feature="assembly_fragment"
misc_feature 1004..2634
/feature="assembly_fragment"
misc_feature 2735..4415
/feature="assembly_fragment"
misc_feature 4516..5785

```

```

/misc_feature /note="assembly_fragment"
5885..7879
/note="assembly_fragment"
7980..9686
/note="assembly_fragment"
9787..12253
/note="assembly_fragment"
12354..15228
/note="assembly_fragment"
15329..17200
/note="assembly_fragment"
17301..20131
/note="assembly_fragment"
20232..22587
/note="assembly_fragment"
22688..25707
/note="assembly_fragment"
25808..28184
/note="assembly_fragment"
28285..31338
/note="assembly_fragment"
31439..34299
/note="assembly_fragment"
34400..38318
/note="assembly_fragment"
38419..42835
/note="assembly_fragment"
42936..45448
/note="assembly_fragment"
45549..48116
/note="assembly_fragment"
48217..52618
/note="assembly_fragment"
52719..56592
/note="assembly_fragment"
56693..59635
/note="assembly_fragment"
59736..63661
/note="assembly_fragment"
63762..68437
/note="assembly_fragment"
68538..71458
/note="assembly_fragment
clone_end:SP6
vector_side:left"
71599..76888
Query Match 58.5%; Score 896; DB 2; Length 175064;
Best Local Similarity 99.9%; Pred. No. 7.3e-229;
Matches 907; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 625 ATAAAGTTGGAGTCTGGAGTGGCTGATACATTAATTTACAGCAGTGGCAATTCCTGGCTGCAC 584
Db 2735 ATAAGTTGGAGTCTGGAGTGGCTGATACATTAATTTACAGCAGTGGCAATTCCTGGCTGCAC 2794

Qy 585 TTCACAAATGATACAACTAAATACAGTCTGTTTATACAGGGAGGCTGATCA 744
Db 2795 TTCACAAATGATACAACTAAATACAGTCTGTTTATACAGGGAGGCTGATCA 2854

Qy 745 ATATAATGAATATAAAGGGGCGTGCCATATCTCTCTGTTTGTGTTTGTGTTT 804
Db 2855 ATATAATGAATATAAAGGGGCGTGGT-CCATA-TGTTCTGTTGTTTGTGTTT 2913

Qy 805 CTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 864
Db 2914 CTTTGTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2973

Qy 865 CTCTCGGTCTCTGAGGAGCTGGGAGCTAGGCTGGGATCTCCAGGCGAGTAGGTC 924
Db 2974 CTCTCGGTCTCTGAGGAGCTGGGAGCTAGGCTGGGATCTCCAGGCGAGTAGGTC 9033

Qy 925 GCCTATCAAAATCAAGTCCAGGTTTGTGGGGGAAACAAAGCAGGCCATTACCCAG 984

```

```

Db 3034 GCCTATCAAAATCAAGTCCAGGTTTGTGGGGGAAACAAAGCAGGCCATTACCCAG 3093
Qy 985 AGGACTGTCCGCTTCCCTCACCACCGCTAGGCTTTTGAAGAGAAACAAAAGACAAGA 1044
Db 3094 AGGACTGTCCGCTTCCCTCACCACCGCTAGGCTTTTGAAGAGAAACAAAAGACAAGA 3153
Qy 1045 CAAAATGATTGGCTCTCTGAGGAGATTACGCTAGAGCTCTCTCTCCCAATCCCTCC 1104
Db 3154 CAAAATGATTGGCTCTCTGAGGAGATTACGCTAGAGCTCTCTCTCCCAATCCCTCC 3213
Qy 1105 CTCGGCTGAGAAACTACAAAGAAACAAAATTCGGAAGAGGATTTAGAGGAAG 1164
Db 3214 CTCGGCTGAGAAACTACAAAGAAACAAAATTCGGAAGAGGATTTAGAGGAAG 3273
Qy 1165 CAAATTCACACTGTGCTTGGCTGGCGGGAACGTTGAGTACAGAGTCTGCGGCGCAGCC 1224
Db 3274 CAAATTCACACTGTGCTTGGCTGGCGGGAACGTTGAGTACAGAGTCTGCGGCGCAGCC 3333
Qy 1225 CCGAGCCACAGCTTCCCGCGGCTTTAGCGCGCGCGGCGCGGCGGGAAGGGGACGC 1284
Db 3334 CCGAGCCACAGCTTCCCGCGGCTTTAGCGCGCGCGGCGCGGCGGGAAGGGGACGC 3393
Qy 1285 AGACCGCGGAGACCTAGACACCTGCTGTACCTCCACCCGCCACCCACCTCCCTCC 1344
Db 3394 AGACCGCGGAGACCTAGACACCTGCTGTACCTCCACCCGCCACCCACCTCCCTCC 3453
Qy 1345 AACTCCCTAGATGTGCTGGCGGCTGAACTGCGGCGCGCTTAAAGGGGCGGCGCGGCT 1404
Db 3454 AACTCCCTAGATGTGCTGGCGGCTGAACTGCGGCGCGCTTAAAGGGGCGGCGGCT 3513
Qy 1405 CCACGTGCTTCTGCTGAGTGAATACATTAACAGAGCGCGGGAACGGGCGGGA 1464
Db 3514 CCACGTGCTTCTGCTGAGTGAATACATTAACAGAGCGCGGGAACGGGCGGGA 3573
Qy 1465 GGAGGAGAGCAGACGCTTTGACCGATAGTAACCTGCGCTGCGTGCAGCCGAATCTAT 1524
Db 3574 GGAGGAGAGCAGACGCTTTGACCGATAGTAACCTGCGCTGCGTGCAGCCGAATCTAT 3633

Qy 1525 AAAAGGAA 1532
Db 3634 AAAAGGAA 3641

RESULT 10
HSA252201 1167 bp DNA linear PRI 10-APR-2001
LOCUS Homo sapiens partial ABC-1 gene for ATP-binding cassette
DEFINITION transporter-1, 5'UTR and promoter region.
ACCESSION AJ252201
VERSION AJ252201.1 GI:12053757
KEYWORDS ABC-1 gene; ATP-binding cassette transporter-1; promoter.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1167)
Porsch-Oezcuernmez,M., Langmann,T., Heimerl,S., Borsukova,H.,
Kaminski,W.E., Drobnik,W., Honer,C., Schumacher,C. and Schmitz,G.
The zinc finger protein 202 (ZNF202) is a transcriptional repressor
of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene
expression and a modulator of cellular lipid efflux
J. Biol. Chem. 276 (15), 12427-12433 (2001)
JOURNAL MEDLINE 21192304
PUBMED 11279031
REFERENCE 2 (bases 1 to 1167)
Porsch-Oezcuernmez,M.K.
AUTHORS Direct Submission
TITLE Submitted (05-JAN-2000) Porsch-Oezcuernmez M.K., Institute for
JOURNAL Clinical Chemistry, University of Regensburg,
Franz-Josef-Strauss-Allee 11, 93042 Regensburg, GERMANY
FEATURES
source
1. .1167
/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/chromosome="9"
/map="9q22-31"
/cell_type="leukocyte"
1..1167
/gene="ABC-1"
promoter
1..942
/gene="ABC-1"
/function="cholesterol efflux regulatory protein"
896..900
/gene="ABC-1"
1148..1167
/gene="ABC-1"
/function="cholesterol efflux regulatory protein"
BASE COUNT 278 a 315 c 327 g 247 t
ORIGIN

Query Match 58.4%; Score 893.4; DB 9; Length 1167;
Best Local Similarity 99.8%; Pred. No. 6.6e-229;
Matches 906; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 626 TAAGTTGGAGCTGGAGTGGCTACATAAATTTACAGACTGCAATTCCTGGCTGCACI 685
DB 1 TAAGTTGGAGCTGGAGTGGCTACATAAATTTACAGACTGCAATTCCTGGCTGCACI 60
QY 686 TCACAAATGTATACAACTAAACAAAGTCTGTTTATACAGAGGAGCTGATCAA 745
DB 61 TCACAAATGTATACAACTAAACAAAGTCTGTTTATACAGAGGAGCTGATCAA 220
QY 746 TATAATGAATTAAGGGGCTGCTCCATATATGTTCTGTTTGTGTTTGTGTTTC 805
DB 121 TATAATGAATTAAGGGGCTGCTCCATATATGTTCTGTTTGTGTTTGTGTTTC 179
QY 806 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTC 865
DB 180 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTC 239
QY 866 TCTGGGTCCTCTGAGGACCTGGGAGCTCAGGCTGAGGCTCAGGCTGAGGCTGAGGCTG 925
DB 240 TCTGGGTCCTCTGAGGACCTGGGAGCTCAGGCTGAGGCTCAGGCTGAGGCTGAGGCTG 299
QY 926 CCTATCAAAATCAAAAGTCCAGGTTGTGGGGGAAACAAAGCAGCCATACCCAGA 985
DB 300 CCTATCAAAATCAAAAGTCCAGGTTGTGGGGGAAACAAAGCAGCCATACCCAGA 359
QY 986 GCACTGTCGGCTTCCCTCACCAGCTGAGCTTGAAGGAAACAAAGCAGCCATACCCAGA 1045
DB 360 GCACTGTCGGCTTCCCTCACCAGCTGAGCTTGAAGGAAACAAAGCAGCCATACCCAGA 429
QY 1046 AAAATGATGGGCTCTGAGGAGATTCAGCTCCTCTCTCCCTCCCAATCCCTCC 1105
DB 420 AAAATGATGGGCTCTGAGGAGATTCAGCTCCTCTCTCCCTCCCAATCCCTCC 479
QY 1106 TCCGGTGAAGAACTAACAAAGGAAAAAATAATTCGGAAGCAGATTTAGAGGAGC 1165
DB 480 TCCGGTGAAGAACTAACAAAGGAAAAAATAATTCGGAAGCAGATTTAGAGGAGC 539
QY 1166 AATTTCCACTGTTGCTTGGCTGCGGAACTGAGCTGAGAGTCTGCGGCGCAGCC 1225
DB 540 AATTTCCACTGTTGCTTGGCTGCGGAACTGAGCTGAGAGTCTGCGGCGCAGCC 599
QY 1226 CGAGCCGAGGCTTCCCGCGCTTTAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1285
DB 600 CGAGCCGAGGCTTCCCGCGCTTTAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 659
QY 1286 GACGCGGAGCCCTTAAGACACTGCTGTACCTCCACCCCAACCCACCCACCTCCCCCA 1345
DB 660 GACGCGGAGCCCTTAAGACACTGCTGTACCTCCACCCCAACCCACCCACCTCCCCCA 719
QY 1346 ACTCCCTAGATGTCTGCTGGCGCTCAAGCTGCGCGCTTAAAGGGGCGGCGGCGGCTC 1405
DB 720 ACTCCCTAGATGTCTGCTGGCGCTCAAGCTGCGCGCTTAAAGGGGCGGCGGCGGCTC 779

```

```

QY 1406 CACGTGCTTCTGCTGAGTGACTGAATACATAAAGAGGCGGCGGCGGCGGAG 1465
DB 780 CACGTGCTTCTGCTGAGTGACTGAATACATAAAGAGGCGGCGGCGGCGGAG 839
QY 1466 GAGGAGACACAGGCTTTGACGATAGTAACCTCTGCGCTCGGTGAGCCGAATCTATA 1525
DB 840 GAGGAGACACAGGCTTTGACGATAGTAACCTCTGCGCTCGGTGAGCCGAATCTATA 899
QY 1526 AAAGGAA 1532
DB 900 AAAGGAA 906

RESULT 11
AF258623S1 1167 bp DNA linear PRI 23-JUN-2000
LOCUS Homo sapiens ATP3 binding cassette transporter 1 (ABCA1) gene,
DEFINITION promoter and exon 1.
ACCESSION AF258623 GI:8677405
VERSION AF258623.2
KEYWORDS
SEGMENT
SOURCE i of 4
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1167)
AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C.,
Aouizerat,B.E., Fielding,C.J. and Kane,J.P.
TITLE Analysis of hABC1 gene 5' end: additional peptide sequence,
promoter region, and four polymorphisms
JOURNAL Biochem. Biophys. Res. Commun. 271 (2000) In press
REFERENCE 2 (bases 224 to 1167)
AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C.,
Aouizerat,B.E., Fielding,C.J. and Kane,J.P.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2000) Cardiovascular Research Institute,
University of California, San Francisco, 505 Parnassus Avenue, San
Francisco, CA 94143-0130, USA
REFERENCE 3 (bases 1 to 1167)
AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C.,
Aouizerat,B.E., Fielding,C.J. and Kane,J.P.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-2000) Cardiovascular Research Institute,
University of California, San Francisco, 505 Parnassus Avenue, San
Francisco, CA 94143-0130, USA
REMARK Sequence update by submitter
COMMENT On Jun 23, 2000 this sequence version replaced gi:7769713.
FEATURES
source
1..1167
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9q31"
224..844
/gene="ABCA1"
845..1147
/gene="ABCA1"
/number=1
BASE COUNT 278 a 313 c 328 g 248 t
ORIGIN

Query Match 58.3%; Score 893.4; DB 9; Length 1167;
Best Local Similarity 99.8%; Pred. No. 1.8e-228;
Matches 905; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 626 TAAGTTGGAGCTGGAGTGGCTACATAAATTTACAGACTGCAATTCCTGGCTGCACI 685
DB 1 TAAGTTGGAGCTGGAGTGGCTACATAAATTTACAGACTGCAATTCCTGGCTGCACI 60
QY 686 TCACAAATGTATACAACTAAACAAAGTCTGTTTATACAGAGGAGCTGATCAA 745
DB 61 TCACAAATGTATACAACTAAACAAAGTCTGTTTATACAGAGGAGCTGATCAA 120
QY 746 TATAATGAATTAAGGGGCTGCTCCATATATGTTCTGTTTGTGTTTGTGTTTC 805
DB 121 TATAATGAATTAAGGGGCTGCTCCATATATGTTCTGTTTGTGTTTGTGTTTC 179
QY 806 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTC 865
DB 180 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTC 239
QY 866 TCTGGGTCCTCTGAGGACCTGGGAGCTCAGGCTGAGGCTCAGGCTGAGGCTGAGGCTG 925
DB 240 TCTGGGTCCTCTGAGGACCTGGGAGCTCAGGCTGAGGCTCAGGCTGAGGCTGAGGCTG 299
QY 926 CCTATCAAAATCAAAAGTCCAGGTTGTGGGGGAAACAAAGCAGCCATACCCAGA 985
DB 300 CCTATCAAAATCAAAAGTCCAGGTTGTGGGGGAAACAAAGCAGCCATACCCAGA 359
QY 986 GCACTGTCGGCTTCCCTCACCAGCTGAGCTTGAAGGAAACAAAGCAGCCATACCCAGA 1045
DB 360 GCACTGTCGGCTTCCCTCACCAGCTGAGCTTGAAGGAAACAAAGCAGCCATACCCAGA 429
QY 1046 AAAATGATGGGCTCTGAGGAGATTCAGCTCCTCTCTCCCTCCCAATCCCTCC 1105
DB 420 AAAATGATGGGCTCTGAGGAGATTCAGCTCCTCTCTCCCTCCCAATCCCTCC 479
QY 1106 TCCGGTGAAGAACTAACAAAGGAAAAAATAATTCGGAAGCAGATTTAGAGGAGC 1165
DB 480 TCCGGTGAAGAACTAACAAAGGAAAAAATAATTCGGAAGCAGATTTAGAGGAGC 539
QY 1166 AATTTCCACTGTTGCTTGGCTGCGGAACTGAGCTGAGAGTCTGCGGCGCAGCC 1225
DB 540 AATTTCCACTGTTGCTTGGCTGCGGAACTGAGCTGAGAGTCTGCGGCGCAGCC 599
QY 1226 CGAGCCGAGGCTTCCCGCGCTTTAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1285
DB 600 CGAGCCGAGGCTTCCCGCGCTTTAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 659
QY 1286 GACGCGGAGCCCTTAAGACACTGCTGTACCTCCACCCCAACCCACCCACCTCCCCCA 1345
DB 660 GACGCGGAGCCCTTAAGACACTGCTGTACCTCCACCCCAACCCACCCACCTCCCCCA 719
QY 1346 ACTCCCTAGATGTCTGCTGGCGCTCAAGCTGCGCGCTTAAAGGGGCGGCGGCGGCTC 1405
DB 720 ACTCCCTAGATGTCTGCTGGCGCTCAAGCTGCGCGCTTAAAGGGGCGGCGGCGGCTC 779

```

2 (bases 1 to 89570)

Biren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Balow, J., Barna, N., Beckerly, R., Bede, F., Boguslavsky, L., Boughgatter, B., Brown, A., Burkett, G., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J., Ferrelira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Kleir, J., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, J., McEwen, P., McGurk, A., McKernan, K., McPheders, R., Meldrum, N., Mewius, L., Morrow, J., Naylor, J., Norman, C.H., O'Connor, J., O'Donnell, P., Olivari, T.M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliou, H., Viel, R., Vo, A., Wu, R., Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6705877.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center -----

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WISR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information -----

Center project name: L2512

Center clone name: L_N10

* NOTE: This record contains 73 individual

* sequencing reads that have not been assembled into

* contigs. Runs of N are used to separate the reads

* and the order in which they appear is completely

* arbitrary. Low-pass sequence sampling is useful for

* identifying clones that may be gene-rich and allows

* overlap relationships among clones to be deduced.

* However, it should not be assumed that this clone

* will be sequenced to completion. In the event that

* the record is updated, the accession number will

* be preserved.

1 871: contig of 871 bp in length

972 971: gap of 100 bp

972 1834: contig of 863 bp in length

1835 -934: gap of 100 bp

1935 2804: contig of 870 bp in length

2805 2904: gap of 100 bp

2905 3745: contig of 841 bp in length

3746 3845: gap of 100 bp

3846 4696: contig of 851 bp in length

4697 4796: gap of 100 bp

4797 5640: contig of 844 bp in length

5641 5740: gap of 100 bp

5741 6540: contig of 800 bp in length

6541 6640: gap of 100 bp

6641 7509: contig of 869 bp in length

7510 7609: gap of 100 bp

7610 8479: contig of 870 bp in length

8480 8579: gap of 100 bp

8580 9430: contig of 851 bp in length

9431 9530: gap of 100 bp

9531 10376: contig of 846 bp in length

10377 10476: gap of 100 bp

10477 11322: contig of 845 bp in length

11323 11422: gap of 100 bp

11423 12302: contig of 880 bp in length

12303 12402: gap of 100 bp

12403 13280: contig of 878 bp in length

13281 13380: gap of 100 bp

13381 13380: gap of 100 bp

REFERENCE
AUTHORS

QY	745	TATAATGAATTAAGAAGGGGCTGGTCCCATATGTTCTCTGTTTGTGTTGTTCTTC	805
Db	121	TATAATGAATTAAGAAGGGGCTGGTCCCATATGTTCTCTGTTTGTGTTGTTCTTC	179
QY	806	TTTTTTGTTTGTGTTGGGCTTCCTCTCTCAATTTATGAGAGAAGCAGTAAAGTTC	865
Db	180	TTTTTTGTTTGTGTTGGGCTTCCTCTCTCAATTTATGAGAGAAGCAGTAAAGTTC	239
QY	866	TCGCGGTCTCTGAGGGACCTGGGAGCTCAGGCTGGGATCTCCCAAGGCATGAGTCG	925
Db	240	TCGCGGTCTCTGAGGGACCTGGGAGCTCAGGCTGGGATCTCCCAAGGCATGAGTCG	299
QY	926	CCATCAAAAATCAAAAGTCCAGGTTTGTGGGGGAAAAACAAGACACGCCATTACCCAGA	985
Db	300	CCATCAAAAATCAAAAGTCCAGGTTTGTGGGGGAAAAACAAGACACGCCATTACCCAGA	359
QY	986	GGACTGTCCGCCCTTCCCTCACCCAGCCTAGGCCCTTGAAGAGAAACAAGACAGAC	1045
Db	360	GGACTGTCCGCCCTTCCCTCACCCAGCCTAGGCCCTTGAAGAGAAACAAGACAGAC	419
QY	1046	AAATGATTGGCTGCTCTGAGGAGATTCAGCTCAGAGCTCTCTCTCCCAATCCCTCC	1105
Db	420	AAATGATTGGCTGCTCTGAGGAGATTCAGCTCAGAGCTCTCTCTCCCAATCCCTCC	479
QY	1106	TCCGGCTGAGGAACCTAACAAAGGAAAAAATAATTCGGGAAGCAGGATTTAGAGGAAGC	1165
Db	480	TCCGGCTGAGGAACCTAACAAAGGAAAAAATAATTCGGGAAGCAGGATTTAGAGGAAGC	539
QY	1166	AAATTCACACTGGTGCCCTTGGCTGCCGGGAAGTGGACTAGAGAGCTCTGGCGCGCAGCCC	1225
Db	540	AAATTCACACTGGTGCCCTTGGCTGCCGGGAAGTGGACTAGAGAGCTCTGGCGCGCAGCCC	599
QY	1226	CGAGCCAGCGCTTCCCGCGCTCTAGCGCGCGGCGCGCGCGGGGGAAGGGGACGCA	1285
Db	600	CGAGCCAGCGCTTCCCGCGCTCTAGCGCGCGGCGCGCGCGGGGGAAGGGGACGCA	659
QY	1286	GACCGCGGACCTAAGACACCTGCTACCTCTCACCCCAACCCCAACCTCCCTCCCA	1345
Db	660	GACCGCGGACCTAAGACACCTGCTACCTCTCACCCCAACCCCAACCTCCCTCCCA	719
QY	1346	ACTCCCTAGATGTTCTGGGGGCTGTAAGCTGCGCCGCTTTAAGGGCGGCGCGCGGCTC	1405
Db	720	ACTCCCTAGATGTTCTGGGGGCTGTAAGCTGCGCCGCTTTAAGGGCGGCGCGCGGCTC	779
QY	1406	CACGTGCTTCTGCTGAGTGACTGAAGTACATAAAGAGAGCGCGGGAAGCGGGCGGGGAG	1465
Db	780	CACGTGCTTCTGCTGAGTGACTGAAGTACATAAAGAGAGCGCGGGAAGCGGGCGGGGAG	839
QY	1466	GAGGAGAGACACAGGCTTTGACCGATAGTAACTCTGCGCTCGGTGCACGCCGAATCTATA	1529
Db	840	GAGGAGAGACACAGGCTTTGACCGATAGTAACTCTGCGCTCGGTGCACGCCGAATCTATA	899
QY	1526	AAAGAA 1532	
Db	900	AAAGAA 906	
RESULT 12			
LOCUS	AC021246	69570 bp	DNA linear HTG 13-JUL
DEFINITION	Homo sapiens clone Rpl1-IN10, LOW-PASS SEQUENCE SAMPLING.		
ACCESSION	AC021246		
VERSION	AC021246.2 GI:9119882		
KEYWORDS	HTG; HTGS_PHASE0.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 69570)		
JOURNAL	Birren,B., Linton,L., Nusbaum,C. and Lander,E.		
	Homo sapiens chromosome, clone Rpl1-IN10		
	Unpublished		

```
* 13381 14241: contig of 861 bp in length
* 14242 14341: gap of 100 bp
* 14342 15196: contig of 855 bp in length
* 15197 15296: gap of 100 bp
* 15297 16123: contig of 827 bp in length
* 16124 16223: gap of 100 bp
* 16224 17072: contig of 849 bp in length
* 17073 17172: gap of 100 bp
* 17173 18041: contig of 869 bp in length
* 18042 18141: gap of 100 bp
* 18142 19009: contig of 868 bp in length
* 19010 19109: gap of 100 bp
* 19110 19966: contig of 857 bp in length
* 19967 20066: gap of 100 bp
* 20067 20921: contig of 855 bp in length
* 20922 21021: gap of 100 bp
* 21022 21865: contig of 844 bp in length
* 21866 21965: gap of 100 bp
* 21966 22832: contig of 867 bp in length
* 22833 22932: gap of 100 bp
* 22933 23780: contig of 848 bp in length
* 23781 23880: gap of 100 bp
* 23881 24733: contig of 853 bp in length
* 24734 24833: gap of 100 bp
* 24834 25670: contig of 837 bp in length
* 25671 25770: gap of 100 bp
* 25771 25621: contig of 851 bp in length
* 25622 26721: gap of 100 bp
* 26722 27576: contig of 855 bp in length
* 27577 27676: gap of 100 bp
* 27677 28532: contig of 856 bp in length
* 28533 28632: gap of 100 bp
* 28633 29492: contig of 880 bp in length
* 29493 29592: gap of 100 bp
* 29593 30455: contig of 863 bp in length
* 30456 30555: gap of 100 bp
* 30556 31410: contig of 855 bp in length
* 31411 31510: gap of 100 bp
* 31511 32368: contig of 858 bp in length
* 32369 32468: gap of 100 bp
* 32469 33312: contig of 844 bp in length
* 33313 33412: gap of 100 bp
* 33413 34268: contig of 856 bp in length
* 34269 34368: gap of 100 bp
* 34369 35204: contig of 836 bp in length
* 35205 35304: gap of 100 bp
* 35305 36156: contig of 852 bp in length
* 36157 36256: gap of 100 bp
* 36257 37128: contig of 872 bp in length
* 37129 37228: gap of 100 bp
* 37229 38083: contig of 855 bp in length
* 38084 38183: gap of 100 bp
* 38184 39031: contig of 848 bp in length
* 39032 39131: gap of 100 bp
* 39132 40006: contig of 875 bp in length
* 40007 40106: gap of 100 bp
* 40107 40967: contig of 861 bp in length
* 40968 41067: gap of 100 bp
* 41068 41913: contig of 846 bp in length
* 41914 42013: gap of 100 bp
* 42014 42824: contig of 811 bp in length
* 42825 42924: gap of 100 bp
* 42925 43776: contig of 852 bp in length
* 43777 43876: gap of 100 bp
* 43877 44752: contig of 876 bp in length
* 44753 44852: gap of 100 bp
* 44853 45724: contig of 872 bp in length
* 45725 45824: gap of 100 bp
* 45825 46643: contig of 819 bp in length
* 46644 46743: gap of 100 bp
* 46744 47599: contig of 856 bp in length
* 47600 47699: gap of 100 bp
* 47700 48551: contig of 852 bp in length
```

```
* 48552 48651: gap of 100 bp
* 48652 49485: contig of 834 bp in length
* 49486 49585: gap of 100 bp
* 49586 50440: contig of 855 bp in length
* 50441 50540: gap of 100 bp
* 50541 51404: contig of 864 bp in length
* 51405 51504: gap of 100 bp
* 51505 52372: contig of 868 bp in length
* 52373 52472: gap of 100 bp
* 52473 53328: contig of 856 bp in length
* 53329 53428: gap of 100 bp
* 53429 54268: contig of 840 bp in length
* 54269 54368: gap of 100 bp
* 54369 55229: contig of 861 bp in length
* 55230 55329: gap of 100 bp
* 55330 56197: contig of 868 bp in length
* 56198 56297: gap of 100 bp
* 56298 57163: contig of 866 bp in length
* 57164 57263: gap of 100 bp
* 57264 58130: contig of 867 bp in length
* 58131 58230: gap of 100 bp
* 58231 59082: contig of 852 bp in length
* 59083 59182: gap of 100 bp
* 59183 60020: contig of 838 bp in length
* 60021 60120: gap of 100 bp
* 60121 60983: contig of 863 bp in length
* 60984 61083: gap of 100 bp
* 61084 61935: contig of 852 bp in length
* 61936 62035: gap of 100 bp
* 62036 62866: contig of 831 bp in length
* 62867 62966: gap of 100 bp
* 62967 63827: contig of 861 bp in length
* 63828 63927: gap of 100 bp
* 63928 64783: contig of 856 bp in length
* 64784 64883: gap of 100 bp
* 64884 65740: contig of 857 bp in length
* 65741 66840: gap of 100 bp
* 66841 66684: contig of 844 bp in length
* 66685 66784: gap of 100 bp
* 66785 67551: contig of 867 bp in length
* 67552 67751: gap of 100 bp

Query Match      26.7%; Score 408.8; DB 2; Length 69570;
Best Local Similarity 99.1%; Pred. No. 3.5e-98;
Matches 432; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1099 CCTCCCTCGGCTGAGAACTAACAAA-GGAAAAAATAATTCGGAAAGCAGGATTTA 1157
DB 41109 CCTCCCTCGGCTGAGAACTAACAAAAGGAAAAAATAATTCGGAAAGCAGGATTTA 41168

QY 1158 GAGGAAGCAAAATTCCTACTGGTGGCTGCCGGGAGCTGGACTAGAGACTTCGGG 1217
DB 41169 GAGGAAGCAAAATTCCTACTGGTGGCTGCCGGGAGCTGGACTAGAGACTTCGGG 41228

QY 1218 CGAGGCCCGAGCCCGAGCGCTTCCCGCGCTTTAGCCCGCGGCGCGGGGGGAAG 1277
DB 41229 CGAGGCCCGAGCCCGAGCGCTTCCCGCGCTTTAGCCCGCGGCGCGGGGGGAAG 41288

QY 1278 GGGAGCAGACCGCGGACCGCTTAAGACACCTGCTGATCCCTCCACCCCGCCACCCCAACC 1337
DB 41289 GGGAGCAGACCGCGGACCGCTTAAGACACCTGCTGATCCCTCCACCCCGCCACCCCAACC 41348

QY 1338 TCCCCCAACT-CCCTAGATGTTCGTGGCGGCTGAACCTGCCCGCTTTAAGGGCGGG 1396
DB 41349 TCCCCCAACTCCCTATATGTGTGCTGGCGGCTGAACCTGCCCGCTTTAAGGGCGGG 41408

QY 1397 CCGCGGCTCCAGTGTCTTCTGCTGAGTGACTAGACTAAACAGACGCGCGGACGG 1456
DB 41409 CCGCGGCTCCAGTGTCTTCTGCTGAGTGACTAAACAGACGCGCGGACGG 41468

QY 1457 GCGGGGAGGAGGAGACAGGCTTTGACCGATAGTACCTCTGCGCTCGGTGCAGCC 1516
DB 41469 GCGGGGAGGAGGAGACAGGCTTTGACCGATAGTACCTCTGCGCTCGGTGCAGCC 41528
```

```

QY 1517 GAATCTATAAAGGAA 1532
      | | | | | | | | | |
DB 41529 GAATCTATAAAGGAA 41544

RESULT 13
AC021246/c
LOCUS AC021246
DEFINITION Homo sapiens clone RP11-1N10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC021246
VERSION 2
KEYWORDS HTG; HTGS-PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 69570)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome, clone RP11-1N10
Unpublished
2 (bases 1 to 69570)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Bouckgaier,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lebecky,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McMan,P., McGurk,A., McKernan,K.,
McPheters,R., Melidrim,J., Meneus,L., Morrow,S., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6705871.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIGR
Web site: http://www-seq.wi.mit.edu
Contact: sequence-submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2512
Center clone name: 1_N10
-----
* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 871: contig of 871 bp in length
*
* 872 971: gap of 100 bp
*
* 972 1834: contig of 863 bp in length
*
* 1835 1934: gap of 100 bp
*
* 1935 2804: contig of 870 bp in length
*
* 2805 2904: gap of 100 bp
*
* 2905 3745: contig of 841 bp in length

```


[illegible]

[illegible]

Search completed: February 4, 2003, 01:14:38
Job time : 5324.17 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comp-gen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 12:46:48 ; Search time 126.452 Seconds
(without alignments)
8245.650 Million cell updates/sec

Title: US-09-596-141C-3_COPY_1181_1643
Perfect score: 463
Sequence: 1 ccttgctcgggaacgtg.....caggcaggggggagcctc 463

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues 4370478

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*

1: /SID22/cgdata/geneseq/geneseqn-embl/NA1980.DAT.*
2: /SID22/cgdata/geneseq/geneseqn-embl/NA1981.DAT.*
3: /SID22/cgdata/geneseq/geneseqn-embl/NA1982.DAT.*
4: /SID22/cgdata/geneseq/geneseqn-embl/NA1983.DAT.*
5: /SID22/cgdata/geneseq/geneseqn-embl/NA1984.DAT.*
6: /SID22/cgdata/geneseq/geneseqn-embl/NA1985.DAT.*
7: /SID22/cgdata/geneseq/geneseqn-embl/NA1987.DAT.*
8: /SID22/cgdata/geneseq/geneseqn-embl/NA1988.DAT.*
9: /SID22/cgdata/geneseq/geneseqn-embl/NA1989.DAT.*
10: /SID22/cgdata/geneseq/geneseqn-embl/NA1990.DAT.*
11: /SID22/cgdata/geneseq/geneseqn-embl/NA1991.DAT.*
12: /SID22/cgdata/geneseq/geneseqn-embl/NA1992.DAT.*
13: /SID22/cgdata/geneseq/geneseqn-embl/NA1993.DAT.*
14: /SID22/cgdata/geneseq/geneseqn-embl/NA1994.DAT.*
15: /SID22/cgdata/geneseq/geneseqn-embl/NA1995.DAT.*
16: /SID22/cgdata/geneseq/geneseqn-embl/NA1996.DAT.*
17: /SID22/cgdata/geneseq/geneseqn-embl/NA1997.DAT.*
18: /SID22/cgdata/geneseq/geneseqn-embl/NA1998.DAT.*
19: /SID22/cgdata/geneseq/geneseqn-embl/NA1999.DAT.*
20: /SID22/cgdata/geneseq/geneseqn-embl/NA2000.DAT.*
21: /SID22/cgdata/geneseq/geneseqn-embl/NA2001A.DAT.*
22: /SID22/cgdata/geneseq/geneseqn-embl/NA2001B.DAT.*
23: /SID22/cgdata/geneseq/geneseqn-embl/NA2001B.DAT.*
24: /SID22/cgdata/geneseq/geneseqn-embl/NA2002.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	463	100.0	1197	24	Human large ATP-binding cassette transporter 1 (hABC1) promoter sequence.
2	463	100.0	1643	22	Human ABC1 gene ex
3	463	100.0	1643	22	Human ABC1 gene ex
4	447	96.5	18399	22	Human ABC1 gene ex
5	446.4	96.4	3231	24	Human ABC1 gene ex
6	372.4	80.4	29.0	24	Human ABC1 gene ex
7	119.4	25.8	227	21	Human ABC1 gene ex
8	98	21.2	7260	22	Human ABC1 gene ex
9	93	21.2	7260	22	Human ABC1 gene ex

ALIGNMENTS

RESULT 1
ABL58400 standard: DNA; 1197 BP.

ABL58400;
30-JUL-2002 (first entry)

Human large ATP-binding cassette transporter 1 (hABC1) promoter sequence.
Human: large ATP-binding cassette transporter 1; ABC1; promoter;
antiarteriosclerotic; gene transfer; transactivator; ds.
Homo sapiens.
OS
WO200183505-A1
PD 08-NOV-2001.
PF 27-APR-2001; 2001WO-US13654.
XX 28-APR-2000; 2000US-0560372.
XX (UYCO) UNIV COLLEGE NEW YORK.
PA Tail AR;
XX WPI; 2002-049334/06.
XX Novel isolated human large ATP-binding cassette transporter 1 promoter
capable of directing transcription of heterologous coding sequence
positioned downstream to it, useful for expressing foreign DNA in host

```

cells -
XX
PS Claim 1: Fig 3; 68pp; English.
XX
CC The invention relates to an isolated human large ATP-binding cassette
CC transporter 1 (ABCI) promoter capable of directing transcription of
CC heterologous coding sequence positioned downstream to it. The ABCI
CC promoter is useful for expressing foreign DNA in a host cell, by
CC introducing into the host cell a gene transfer vector comprising the
CC promoter operably linked to a foreign DNA encoding a desired polypeptide
CC or RNA, where the foreign DNA is expressed. The gene transfer can be
CC introduced into the host cell by adenovirus infection, liposome-mediated
CC transfer, topical application to the cell or microinjection. The gene
CC transfer vector encodes and expresses a reporter molecule. The method
CC further involves introducing into the cell a gene transfer vector
CC comprising a nucleic acid segment encoding a transactivator protein
CC capable of upregulating the ABCI promoter, or contacting the cell with
CC the transactivator protein, or an agonist of the transactivator protein.
CC Modulators of human ABCI gene expression are useful for treating
CC atherosclerosis. The present sequence represents the ABCI promoter.
XX
SQ Sequence 1197 BP; 284 A; 314 C; 328 G; 271 T; 0 other;

Query Match 100.0%; Score 463; DB 24; Length 1197;
Best Local Similarity 100.0%; Pred. No. 7.8e-108;
Matches 463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTGGCTGCCGGGAACGTGGACTAGAGACTCTGCGGCGACGCCCGAGCCGAGCGCTTC 60
XX |||||
DB 719 CCTTGGCTGCCGGGAACGTGGACTAGAGACTCTGCGGCGACGCCCGAGCCGAGCGCTTC 778
QY 61 CGCGCGCTTTAGGCGCGCGGCCCGCGGGGAAAGGGGACGACGACGCGCGGACCGCTAA 120
DB |||||
DB 779 CGCGCGCTTTAGGCGCGCGGCCCGCGGGGAAAGGGGACGACGACGCGCGGACCGCTAA 838
QY 121 GACACCTGCTGACCTCCACGCCACCCACCCACCCACCTCCGCCACCTCCCTAGATGTG 180
DB |||||
DB 839 GACACCTGCTGACCTCCACGCCACCCACCCACCCACCTCCGCCACCTCCCTAGATGTG 898
QY 181 CGTGGCGCTGAACCTGCGCGCTTTAAGGGCGGGCGGCCCGGCTCCACGTGCTTCTGCT 240
DB |||||
DB 899 CGTGGCGCTGAACCTGCGCGCTTTAAGGGCGGGCGGCCCGGCTCCACGTGCTTCTGCT 958
QY 241 GAGTGACCTGACCTACATACAGAGAGCGCGGACGCGGGCGGGGAGGAGGAGACAG 300
DB |||||
DB 959 GAGTGACCTGACCTACATACAGAGAGCGCGGACGCGGGGAGGAGGAGGAGACAG 1018
QY 301 CTTTGACCGGATAGTAACTCTGCGCTCGGTGCGAGCGCGGATCTATAAAGGAACTAGTCC 360
DB |||||
DB 1019 CTTTGACCGGATAGTAACTCTGCGCTCGGTGCGAGCGCGGATCTATAAAGGAACTAGTCC 1078
QY 361 GGCATAAACCCTGTAATTCGAGCGGAGAGTGTAGTGGGCGGGGACCCGAGAGCGGAGCC 420
DB |||||
DB 1079 GGCATAAACCCTGTAATTCGAGCGGAGAGTGTAGTGGGCGGGGACCCGAGAGCGGAGCC 1138
QY 421 GACCCCTCTCTCCCGGCTGCGGCGAGGCGAGGCGGGGAGCTC 463
DB |||||
DB 1139 GACCCCTCTCTCCCGGCTGCGGCGAGGCGAGGCGGGGAGCTC 1181

RESULT 2
AAF24681
ID AAF24681 standard; DNA; 1643 BP.
XX
AC AAF24681;
XX
DT 20-APR-2001 (first entry)
XX
DE Nucleotide sequence of the 5' flanking region of the human ABCI gene.
XX
KW Human; adenosine triphosphate binding cassette protein 1; ABCI;
KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
KW chromosome 9q22-q31; heart disease; hypercholesterolemia;

```

```

KW atherosclerosis; cholesterol transport; ss.
XX Homo sapiens.
XX
PN WC2000078972-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WC-US16765.
XX
PR 18-JUN-1999; 99US-0140264.
PR 14-SEP-1999; 99US-0153872.
PR 19-NOV-1999; 99US-0166573.
XX
XX (CVTH-) CV THERAPEUTICS INC.
XX
XX Lawn RM, Wade D, Garvin M;
PI WPI; 2001-137812/14.
XX
PT Adenosine triphosphate (ATP) binding cassette (ABC) polynucleotide,
PT useful for the development of agents for the treatment of heart disease
PT and other disorders associated with hypercholesterolemia and
PT atherosclerosis.
XX
PS Claim 1: Page 143-144; 215pp; English.
XX
XX The present sequence represents the 5' flanking region of the human
XX adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABCI
XX resides in cell membranes and utilises ATP hydrolysis to transport a wide
XX variety of substrates across the plasma membrane. ABCI is a pivotal
XX protein in the apolipoprotein-mediated mobilisation of intracellular
XX cholesterol stores. ABCI is defective in Tangier disease, a genetic
XX disorder characterised by abnormal HDL-cholesterol metabolism. The ABCI
XX gene is localised to chromosome 9q22-q31. The ABCI genes and proteins
XX are useful for developing pharmaceutical agents for the treatment of
XX heart disease and other disorders associated with hypercholesterolemia
XX and atherosclerosis. The genes are useful for developing screening assays
XX to screen for compounds that regulate the expression of genes associated
XX with cholesterol transport. The genes and proteins are also useful for
XX are also useful as diagnostic indicators of cardiovascular disease and
XX other disorders associated with hypercholesterolemia.
XX
SQ Sequence 1643 BP; 370 A; 413 C; 457 G; 403 T; 0 other;

```

```

Query Match 100.0%; Score 463; DB 22; Length 1643;
Best Local Similarity 100.0%; Pred. No. 8.2e-108;
Matches 463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTGGCTGCCGGGAACGTGGACTAGAGAGTCTGCGGCGACGCCCGAGCCGAGCGCTTC 60
DB |||||
DB 1181 CCTTGGCTGCCGGGAACGTGGACTAGAGAGTCTGCGGCGACGCCCGAGCCGAGCGCTTC 1240
QY 61 CGCGCGCTTTAGGCGCGCGGCCCGCGGGGAAAGGGGACGACGACGCGGAGCGCTAA 120
DB |||||
DB 1241 CGCGCGCTTTAGGCGCGCGGCCCGCGGGGAAAGGGGACGACGACGCGGAGCGCTAA 1300
QY 121 GACACCTGCTGTACCTCCACGCCACCCACCCACCTCCGCCACCTCCCTAGATGTGT 180
DB |||||
DB 1301 GACACCTGCTGTACCTCCACGCCACCCACCCACCTCCGCCACCTCCCTAGATGTGT 1360
QY 181 CGTGGCGCTGAACCTGCGCGCTTTAAGGGCGGGCGGCCCGGCTCCACGTGCTTCTGCT 240
DB |||||
DB 1361 CGTGGCGCTGAACCTGCGCGCTTTAAGGGCGGGCGGCCCGGCTCCACGTGCTTCTGCT 1420
QY 241 GAGTGACCTGACCTACATACAGAGAGCGCGGAAACGCGGCGGAGGAGGAGACAG 300
DB |||||
DB 1421 GAGTGACCTGACCTACATACAGAGAGCGCGGAAACGCGGCGGAGGAGGAGACAG 1480
QY 301 CTTTGACCGGATAGTAACTCTGCGCTCGGTGCGAGCGCGGATCTATAAAGGAACTAGTCC 360
DB |||||
DB 1481 CTTTGACCGGATAGTAACTCTGCGCTCGGTGCGAGCGGATCTATAAAGGAACTAGTCC 1540

```

QY 361 GCACAAACCCCGTAATTGCGAGCGAGAGTGGTGGGGCGGACCCCGAGAGCCGAGCC 420
 |||||
 Db 1541 GSCAAACCCCGTAATTGCGAGCGAGAGTGGTGGGGCGGACCCCGAGAGCCGAGCC 1600
 |||||
 QY 421 GACCCCTTCTCTCCCGGCTCGCGGACGAGCGGCGGGAGGCTC 463
 |||||
 Db 1601 GACCCCTTCTCTCCCGGCTCGCGGACGAGCGGCGGGAGGCTC 1643
 |||||

RESULT 3

AAF24703
 ID AAF24703 standard; DNA; 1643 BP.
 XX AC
 XX AAF24703;
 DI 20-APR-2001 (first entry)
 XX DE Nucleotide sequence of the 5' flanking region of the human ABC1 gene.
 XX KW Human; adenosine triphosphate binding cassette protein 1; ABC1;
 KW apolipoprotein-mediated mobilisation; cholesterol; Tangier disease;
 KW chromosome 9q22-9q31; heart disease; hypercholesterolemia;
 KW atherosclerosis; cholesterol transport; SS.
 XX OS Homo sapiens.
 XX PN WO200078971-A2.
 XX PD 28-DEC-2000.
 XX PF 16-JUN-2000; 2000WO-US16591.
 XX PR 18-JUN-1999; 99US-0140264.
 XX PR 14-SEP-1999; 99US-0153872.
 XX PR 19-NOV-1999; 99US-0166573.
 XX PA (CVTH-) CV THERAPEUTICS INC.
 XX PA (UNIW) UNIV WASHINGTON.
 XX PI Laws RM, Wade D, Oram JF, Garvin M;
 XX WPI; 2001-13781/14.

XX ADenosine triphosphate (ATP) binding cassette protein (ABC) 1
 XX polynucleotides and polypeptides, useful for treatment of heart disease
 XX and other disorders associated with hypercholesterolemia and
 XX atherosclerosis -
 XX
 XX Disclosure: Page 138-139; 211pp; English.
 XX
 XX The present sequence represents the 5' flanking region of the human
 XX adenosine triphosphate (ATP) binding cassette protein (ABC) 1 gene. ABC1
 XX resides in cell membranes and utilizes ATP hydrolysis to transport a wide
 XX variety of substrates across the plasma membrane. ABC1 is a pivotal
 XX protein in the apolipoprotein-mediated mobilisation of intracellular
 XX cholesterol stores. ABC1 is defective in Tangier disease, a genetic
 XX disorder characterised by abnormal HDL-cholesterol metabolism. The ABC1
 XX gene is localised to chromosome 9q22-9q31. The ABC1 genes and proteins
 XX are useful for developing pharmaceutical agents for the treatment of
 XX heart disease and other disorders associated with hypercholesterolemia
 XX and atherosclerosis. The genes are useful for developing screening assays
 XX to screen for compounds that regulate the expression of genes associated
 XX with cholesterol transport. The genes and proteins are also useful for
 XX are also useful as diagnostic indicators of cardiovascular disease and
 XX other disorders associated with hypercholesterolemia.

XX Sequence 1643 BP; 370 A; 413 C; 457 G; 403 T; 0 other;
 Query Match:
 Best Local Similarity 100.0%; Score 463; DB 22; Length 1643;
 Matches 463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CCTTGGCTCCCGGACGTTGGACTAGAGACTCTGGCGGCGACCCCGAGCCGCGCTTC 60

Db 1181 CTTTSGCTGCCGGACGTTGGACIAGAGAGTCTCGCGCGACCCCGAGCCGCTTC 1243
 |||||
 QY 61 CGCGCGCTTTAGCGCGCGGCGCGCGGCGGGAAGGGAGCGACAGCCGACCCCTAA 120
 |||||
 Db 1241 CGCGCGCTTTAGCGCGCGGCGCGCGGCGGGAAGGGAGCGACAGCCGCGACCCCTAA 1300
 |||||
 QY 121 GACACCTGCTGTACCTCCACCCCGACCCCGACCTCCCGCCCACTCCCTAGATGTGT 180
 |||||
 Db 1301 GACACCTGCTGTACCTCCACCCCGACCCCGACCTCCCGCCCACTCCCTAGATGTGT 1360
 |||||
 QY 181 CTTGSGCGGCTGAACGTCGCCCGGTTTAAGGGCGGCGCGCGGCTCCACGTGTTTGTGCT 240
 |||||
 Db 1361 CTTGSGCGGCTGAACGTCGCCCGGTTTAAGGGCGGCGCGGCTCCACGTGTTTGTGCT 1420
 |||||
 QY 241 GAGTGACTGAACCTAGATAAAGAGAGCGCGGGAACCGGGCGGAGGAGGAGACACAGG 300
 |||||
 Db 1421 GAGTGACTGAACCTAGATAAAGAGAGCGCGGGAACCGGGCGGAGGAGGAGACACAGG 1480
 |||||
 QY 301 CTTTGACCCGATAGTAACCTCTCGGTCGCGTCAGCGCGGAATCTATAAAGGAACCTAGTCCC 360
 |||||
 Db 1481 CTTTGACCGAATAGTAACCTCTCGGTCGCGTCAGCGCGGAATCTATAAAGGAACCTAGTCCC 1540
 |||||
 QY 361 GGCACAAACCCCGTAATTGCGAGCGAGAGTGAATGGGCGCGGACCCCGAGAGCCGAGCC 420
 |||||
 Db 1541 GGCACAAACCCCGTAATTGCGAGCGAGAGTGAATGGGCGCGGACCCCGAGAGCCGAGCC 1600
 |||||
 QY 421 GACCCCTTCTCTCCCGGCTCGCGGAGGCGAGCGGGGAGGCTC 463
 |||||
 Db 1601 GACCCCTTCTCTCCCGGCTCGCGGAGGCGAGCGGGGAGGCTC 1643
 |||||

RESULT 4

AAF92831
 ID AAF92831 standard; DNA; 183999 BP.
 XX AC
 XX AAF92831;
 DI 17-MAY-2001 (first entry)
 XX DE Human ABC1 genomic DNA.
 XX High density lipoprotein-cholesterol; HDL-C; cardiovascular; ABC1; ds.
 XX Homo sapiens.
 XX WO200115676-A2.
 XX PD 08-MAR-2001.
 XX PF 01-SEP-2000; 2000WO-IB01492.
 XX PR 01-SEP-1999; 99US-0151977.
 XX PR 15-MAR-2000; 2000US-0526193.
 XX PR 23-JUN-2000; 2000US-0213958.
 XX (UYBR-) UNIV BRITISH COLUMBIA.
 XX (XENO-) XENON GENETICS INC.
 XX Bayden MR, Brooks-Wilson AR, Pimstone SN, Clee SM;
 XX WPI; 2001-244356/25.
 XX Treating a lower than normal high density lipoprotein-cholesterol
 XX (HDL-C) level, a higher than normal triglyceride level, or a
 XX cardiovascular disease, by administering a compound that modulates LXR-
 XX or RXR-mediated transcriptional activity -
 XX Claim 8; Fig 1; 317pp; English.
 XX The present invention relates to a method for treating a patient
 XX diagnosed as having a lower than normal high density
 XX lipoprotein-cholesterol (HDL-C) level, a higher than normal

CC triglyceride level, or a cardiovascular disease, involving
CC administering a compound that modulates LXR- α or RXR-mediated
CC transcriptional activity or ABC1 expression or activity.
CC The LXR gene product may be used in an assay to identify
CC compounds useful for the treatment of a disease or condition selected a
CC lower than normal HDL cholesterol level, a higher than normal
CC triglyceride level, and a cardiovascular disease.

XX
SQ Sequence 183999 BP; 49549 A; 37944 C; 41170 G; 54950 T; 386 other;

Query Match 96.5%; Score 447; DB 22; Length 183999;
Best Local Similarity 98.7%; Pred. No. 2.1e-103;
Matches 462; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY - CCTTGGCTGCGGAGCGTGGACTAGAGTCTGGCGGAGCCCGGAGCCGAGCGCTTC 60
|||||
Db 28413 CCTTGGCTGCGGAGCGTGGACTAGAGTCTGGCGGAGCCCGGAGCCGAGCGCTTC 28472
|||||
QY 61 CCCTGGCTCTTAGCGCGCGCGCGCGCGGCGGGAAGGGGACGACGCGCGACCCCTAA 120
|||||
Db 28473 CCCTGGCTCTTAGCGCGCGCGCGCGGCGGGAAGGGGACGACGCGCGACCCCTAA 28532
|||||

QY 121 GACACCTGCTGTACCTCCCA-----CCCCACCCCGACCCACCTCCCGCCCTCCCTAGA 175
|||||
Db 28533 GACACCTGCTGTACCTCCCA-----CCCCACCCCGACCCACCTCCCGCCCTCCCTAGA 28592
|||||

QY 176 TGTGCTGCGGCGGCTGAAGTCTGCGGCTTAAGGGCGCGGCGCGCGCTCCAGTGTCTT 235
|||||
Db 28593 TGTGCTGCGGCGGCTGAAGTCTGCGGCTTAAGGGCGCGGCGCGCGCTCCAGTGTCTT 28652
|||||

QY 236 CTCCTGAGTGAAGTGAATACATAAAGAGCGCGGGAAGGGGAGGAGGAGGAGC 295
|||||
Db 28653 CTCCTGAGTGAAGTGAATACATAAAGAGCGCGGGAAGGGGAGGAGGAGGAGC 28712
|||||

QY 296 ACAGGCTTTGACCGATAGTAACCTCTCCGCTCGGTCGAGCCGGAATCTATATAAGGAACTA 355
|||||
Db 28713 ACAGGCTTTGACCGATAGTAACCTCTCCGCTCGGTCGAGCCGGAATCTATATAAGGAACTA 28772
|||||

QY 356 GTCCCGGCAAAACCCCGTAATTCGAGCGAGAGTGAAGGGCGCGGACCCCGAGAGCC 415
|||||
Db 28773 GTCCCGGCAAAACCCCGTAATTCGAGCGAGAGTGAAGGGCGCGGACCCCGAGAGCC 28832
|||||

QY 416 GAGCGACCCCTTCTCTCCCGGCTGCGGAGGCGCAGGCGGGGAGCTC 463
|||||
Db 28833 GAGCGACCCCTTCTCTCCCGGCTGCGGAGGCGCAGGCGGGGAGGCTC 28890
|||||

RESULT 5

AAD37265
ID AAD37265 standard; DNA; 3231 BP.

XX
AC AAD37265;

XX
DT 21-AUG-2002 (first entry)

XX
DE Human ABC1 transcription regulatory DNA #1.

XX
KW Human; ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
KW cholesterol metabolism; hypercholesterolaemia; antisense therapy; ds.

XX
OS Homo sapiens.

XX
PN W0200183745-A2.

XX
PD 08-NOV-2001.

XX
PF 02-MAY-2001; 2001WO-EP05488.

XX
PR 02-MAY-2000; 2000US-201280P.

XX
PA (AVET) AVENTIS PHARMA SA.

XX
PI Rosier-Montus K, Prades C, Lemoine C, Naudin L, Deneffe P;

PI Brewer B, Duvenger N, Remaley A, Santamarina-Fojo S;

XX
DR WPI; 2002-154404/20.

XX
PT Isolated nucleic acid useful for modifying the ATP-binding cassette 1
PT (ABC1) and screening for candidate modulatory compounds or substances

XX
PS Claim 1; Page 130-131; 152pp; English.

XX
CC The invention relates a nucleic acid which is capable of regulating the
CC transcription of human ATP-binding cassette 1 (ABC1) gene, which is a
CC casual gene for pathologies linked to a dysfunctioning of cholesterol
CC metabolism, including diseases such as atherosclerosis. Polynucleotides
CC of the invention are used to screen candidate molecules or substances
CC that are capable of modulating the transcription of the ABC1 gene. They
CC are used in antisense therapy. Compositions comprising sequences of the
CC invention are used to treat hypercholesterolaemia and atherosclerosis.
CC The present sequence is human ABC1 transcription regulating DNA.

XX
SQ Sequence 3231 BP; 809 A; 773 C; 876 G; 773 T; 0 other;

Query Match 96.4%; Score 446.4; DB 24; Length 3231;
Best Local Similarity 98.7%; Pred. No. 1.5e-103;
Matches 462; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY 1 CCTTGGCTGCGGAGCGTGGACTAGAGTCTGGCGGAGCCCGGAGCCGAGCGCTTC 60
|||||
Db 2517 CCTTGGCTGCGGAGCGTGGACTAGAGTCTGGCGGAGCCCGGAGCCGAGCGCTTC 2576
|||||

QY 61 CCCTGGCTCTTAGCGCGCGCGCGCGGCGGGAAGGGGACGACGCGCGACCCCTAA 120
|||||
Db 2577 CCCTGGCTCTTAGCGCGCGCGCGCGGCGGGAAGGGGACGACGCGCGACCCCTAA 2636
|||||

QY 121 GACACCTGCTGTACCTCCCA-----CCCCACCCCGACCCACCTCCCGCCCTCCCTAGA 175
|||||
Db 2637 GACACCTGCTGTACCTCCCA-----CCCCACCCCGACCCACCTCCCGCCCTCCCTAGA 2696
|||||

QY 176 TGTGCTGCGGCGGCTGAAGTCTGCGGCTTAAGGGCGCGGACCCCGCTCCAGTGTCTT 235
|||||
Db 2697 TGTGCTGCGGCGGCTGAAGTCTGCGGCTTAAGGGCGCGGACCCCGCTCCAGTGTCTT 2756
|||||

QY 236 CTCCTGAGTGAAGTGAATACATAAAGAGCGCGGGAAGGGGAGGAGGAGGAGC 295
|||||
Db 2757 CTCCTGAGTGAAGTGAATACATAAAGAGCGCGGGAAGGGGAGGAGGAGGAGC 2816
|||||

QY 296 ACAGGCTTTGACCGATAGTAACCTCTCCGCTCGGTCGAGCCGGAATCTATATAAGGAACTA 355
|||||
Db 2817 ACAGGCTTTGACCGATAGTAACCTCTCCGCTCGGTCGAGCCGGAATCTATATAAGGAACTA 2876
|||||

QY 356 GTCCCGGCAAAACCCCGTAATTCGAGCGAGAGTGAAGGGCGCGGACCCCGAGAGCC 415
|||||
Db 2877 GTCCCGGCAAAACCCCGTAATTCGAGCGAGAGTGAAGGGCGCGGACCCCGAGAGCC 2936
|||||

QY 416 GAGCGACCCCTTCTCTCCCGGCTGCGGAGGCGCAGGCGGGGAGCTC 463
|||||
Db 2937 GAGCGACCCCTTCTCTCCCGGCTGCGGAGGCGCAGGCGGGGAGGCTC 2984
|||||

RESULT 6

AAD37267
ID AAD37267 standard; DNA; 2910 BP.

XX
AC AAD37267;

XX
DT 21-AUG-2002 (first entry)

XX
DE Human ABC1 transcription regulatory DNA #3.

XX
KW Human; ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
KW cholesterol metabolism; hypercholesterolaemia; antisense therapy; ds.

XX
OS Homo sapiens.

XX	Key	Location/Qualifiers	protein_bind	/bound_moiety=
FT	protein_bind	12..23	protein_bind	1622..1627
FT		/*tag= a		/*tag= z
FT	protein_bind	/bound_moiety= "LMO2COM/MYCD"	protein_bind	/bound_moiety= "PPAR"
FT		97..107		1632..1637
FT		/*tag= b		/*tag= aa
FT	protein_bind	/bound_moiety= "DeltaEFl"	protein_bind	/bound_moiety= "PPAR"
FT		110..125		1685..1698
FT		/*tag= c		/*tag= ab
FT	protein_bind	/bound_moiety= "S8/NKX2.5"	protein_bind	/bound_moiety= "USF/NMYC/MYCMAX"
FT		196..211		1787..1797
FT		/*tag= d		/*tag= ac
FT	protein_bind	/bound_moiety= "S8"	protein_bind	/bound_moiety= "DeltaEFl"
FT		228..237		1809..1819
FT		/*tag= e		/*tag= ad
FT	protein_bind	/bound_moiety= "GATA"	protein_bind	/bound_moiety= "DeltaEFl"
FT		399..410		1822..1833
FT		/*tag= f		/*tag= ae
FT	protein_bind	/bound_moiety= "IK2"	protein_bind	/bound_moiety= "SRY"
FT		412..420		1840..1850
FT		/*tag= g		/*tag= af
FT	protein_bind	/bound_moiety= "LYF1"	protein_bind	/bound_moiety= "AP1"
FT		528..539		1942..1956
FT		/*tag= h		/*tag= ag
FT	protein_bind	/bound_moiety= "LMO2COM/MYOD/DeltaEFl"	protein_bind	/bound_moiety= "HNF3beta"
FT		549..556		1978..1985
FT		/*tag= i		/*tag= ah
FT	protein_bind	/bound_moiety= "LYF1"	protein_bind	/bound_moiety= "NKX2.5"
FT		558..568		2008..2016
FT		/*tag= j		/*tag= ai
FT	protein_bind	/bound_moiety= "DeltaEFl"	protein_bind	/bound_moiety= "PPAR/NKX2.5/PPAR"
FT		590..596		2019..2024
FT		/*tag= k		/*tag= aj
FT	protein_bind	/bound_moiety= "LXX2.5"	protein_bind	/bound_moiety= "PPAR/NKX2.5/PPAR"
FT		608..620		2051..2059
FT		/*tag= l		/*tag= ak
FT	protein_bind	/bound_moiety= "NFY/CAAT"	protein_bind	/bound_moiety= "GATA"
FT		708..715		2104..2111
FT		/*tag= m		/*tag= al
FT	protein_bind	/bound_moiety= "MZFl"	protein_bind	/bound_moiety= "SOX5"
FT		723..730		2114..2152
FT		/*tag= n		/*tag= am
FT	protein_bind	/bound_moiety= "MZFl"	protein_bind	/bound_moiety= "SYR/HFH/HNF3beta"
FT		771..785		2221..2228
FT		/*tag= o		/*tag= an
FT	protein_bind	/bound_moiety= "HFH2/SRY/EV11"	protein_bind	/bound_moiety= "MZFl"
FT		803..812		2234..2249
FT		/*tag= p		/*tag= ao
FT	protein_bind	/bound_moiety= "CREBFl/VSp"	protein_bind	/bound_moiety= "IK2/NFKappaB/CREL"
FT		831..837		2259..2272
FT		/*tag= r		/*tag= ap
FT	protein_bind	/bound_moiety= "NKX2.5"	protein_bind	/bound_moiety= "LMO2COM/GATA"
FT		1076..1089		2289..2306
FT		/*tag= s		/*tag= aq
FT	protein_bind	/bound_moiety= "GATA"	protein_bind	/bound_moiety= "MZFl/SRY"
FT		1173..1188		2313..2318
FT		/*tag= t		/*tag= ar
FT	protein_bind	/bound_moiety= "LXRalpha/DeltaEFl"	protein_bind	/bound_moiety= "PPAR"
FT		1189..1209		2321..2326
FT		/*tag= u		/*tag= as
FT	protein_bind	/bound_moiety= "DeltaEFl/LYF1/IK2"	protein_bind	/bound_moiety= "PPAR"
FT		1483..1491		2335..2342
FT		/*tag= v		/*tag= at
FT	protein_bind	/bound_moiety= "AP4"	protein_bind	/bound_moiety= "MZFl"
FT		1498..1514		2361..2384
FT		/*tag= w		/*tag= au
FT	protein_bind	/bound_moiety= "LMO2-COM/MYOD/DeltaEFl/BA7"	protein_bind	/bound_moiety= "HNF3beta/SRY/EV11"
FT		1524..1545		2426..2433
FT		/*tag= x		/*tag= av
FT	protein_bind	/bound_moiety= "ZID/DeltaEFl"	protein_bind	/bound_moiety= "MZFl"
FT		1597..1607		2442..2451
FT		/*tag= y		/*tag= aw
FT				/bound_moiety= "AP4"

```
protein_bind 2455..2466
/*tag= ax
/round_moiety= "SRV"
protein_bind 2491..2498
/*tag= ay
/round_moiety= "STAT"
protein_bind 2524..2534
/*tag= az
/round_moiety= "STAT/PPAR"
protein_bind 2536..2541
/*tag= ba
/round_moiety= "PPAR"
protein_bind 2589..2600
/*tag= bb
/round_moiety= "AP2"
protein_bind 2510..2617
/*tag= bc
/round_moiety= "M2F1"
protein_bind 2634..2648
/*tag= bd
/round_moiety= "LMC2COK/XYOD/E47"
protein_bind 2657..2672
/*tag= be
/round_moiety= "RREB1"
protein_bind 2680..2699
/*tag= bf
/round_moiety= "M2F1/CMYB"
protein_bind 2728..2740
/*tag= bg
/round_moiety= "SP1/GC"
protein_bind 2743..2757
/*tag= bh
/round_moiety= "USF/NMYC/ARNT"
protein_bind 2759..2773
/*tag= bi
/round_moiety= "NFE2A1"
protein_bind 2774..2787
/*tag= bj
/round_moiety= "XFD-/HFH"
protein_bind 2794..2806
/*tag= bk
/round_moiety= "GC/SP1/MZF1"

Query Match 80.4%; Score 372.4; DB 24; Length 2510;
Best Local Similarity 98.5%; Pred. No. 7.7e-85;
Matches 388; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY 1 COTTGGCTGCCGGGAACGTGGACTAGAGAGTCTGCGGCGCAGCCCGAGCCCGAGCGCTTC 60
|||||
DB 2517 COTTGGCTGCCGGGAACGTGGACTAGAGAGTCTGCGGCGCAGCCCGAGCGCGCTTC 2576
|||||

QY 61 CGGCGCTTTAGCGCGGGGCGCGCGCGGGAAGGGGAGCGAGCGCGGAGCCCTAA 120
|||||
DB 2577 CGGCGCTTTAGCGCGGGGCGCGCGCGGGAAGGGGAGCGAGCGCGGAGCCCTAA 2636
|||||

QY 121 GACACTGCTGTACCTCCACCCCAACCCACCC-----ACCTCCCCCACTCCCTAGA 175
|||||
DB 2637 GACACTGCTGTACCTCCACCCCAACCCCAACCCCACTCCCTCCCACTCCCTAGA 2696
|||||

QY 176 TGTGTGTTGGGGGGCTGAACGTCGCGCGGTTTAAAGGGGGGGCGCGGTCACGTGCTT 235
|||||
DB 2697 TGTGTGTTGGGGGGCTGAACGTCGCGCGGTTTAAAGGGGGGGCGCGGTCACGTGCTT 2756
|||||

QY 236 CTGCTGAGTGAAGTACATACAGAGGCGGGAAGGGGCGGGGAGGAGGAGGAGC 295
|||||
DB 2757 CTGCTGAGTGAAGTACATACAGAGGCGGGAAGGGGCGGGGAGGAGGAGGAGC 2816
|||||

QY 296 ACAGGCTTTGACGATAGTAAGCTCTGCGCTCGGTGACCGCAATCTATAAAGGAACTA 355
|||||
DB 2817 ACAGGCTTTGACGATAGTAAGCTCTGCGCTCGGTGACCGCAATCTATAAAGGAACTA 2876
|||||

QY 356 GTCCCCGGCAAAACCCCGTAATTGGAGCGAGAG 389
|||||
```

```
Db 2877 GTCCGCGCAAAACCCCGTAATTGGAGCGAGAG 2910

RESULT 7
AAC09615
ID AAC09615 standard; cDNA: 227 BP.
XX AAC09615;
XX
XX 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 13690.
DE
XX Humar; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX
XX BP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 13690: 71pp + CD-ROM; English.
XX
XX The present sequence is one of a large number of 5' ESTs derived from
XX mRNAs encoding secreted proteins. No ORF has yet been conclusively
XX identified within the present sequence. The 5' ESTs were prepared from
XX total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
XX sequences usually correspond mainly to the 3' untranslated region (UTR)
XX of the mRNA because they are often obtained from oligo-dT primed cDNA
XX libraries. Such ESTs are not well suited for isolating cDNA sequences
XX derived from the 5' ends of mRNAs and even in those cases where longer
XX cDNA sequences have been obtained, the full 5' UTR is rarely included.
XX 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
XX used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
XX in diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX They are used to obtain upstream regulatory sequences and to design
XX expression and secretion vectors.
XX
XX Sequence 227 BP; 44 A; 65 C; 73 G; 45 T; 0 other;
```

```
Query Match 25.8%; Score 119.4; DB 21; Length 227;
Best Local Similarity 99.2%; Pred. No. 5.3e-21;
Matches 120; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 343 ATAAAGGAAGTACTAGTCCCGGCAAAACCCCGTAATTGGAGCGAGAGTCACTGGGCGGG 402
|||||
DB 1 ATAAAGGAAGTACTAGTCCCGGCAAAACCCCGTAATTGGAGCGAGAGTCACTGGGCGGG 60
|||||

QY 403 GACCCGCGAGCGGAGCGGAGCCCTCTCTCCCGGCTCGCGGAGGCGGCGGAGCT 462
|||||
DB 61 GACCCGCGAGCGGAGCGGAGCCCTCTCTCCCGGCTCGCGGAGGCGGCGGAGCT 120
|||||

QY 463 C 463
DB 121 C 121
```

RESULT 8


```

AAD21326
CD AAD21326 standard; DNA; 7260 BP.
XX
AC AAD21326;
XX
DT 28-JAN-2002 (first entry)
XX
DE Human ATP binding cassette transporter 1 (ABCI) gene.
XX
KW Human: ATP binding cassette transporter 1; ABC1; human; lipid disorder;
KW dermatological; atherosclerosis; cardiovascular; coronary heart disease;
KW psoriasis; lipid disorder; antibacterial; septic shock; gene therapy;
KW immunosuppressive; lupus erythematosus; rheumatoid arthritis; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FI CDS 321..7106
FT CDS /*tag= a
FT FT /*product= "Human ABC1 protein"
FT FT
XX
XX EPI136552-A1.
XX
XX 26-SEP-2001.
XX
XX 20-MAR-2000; 2000EP-0105820.
XX
XX 20-MAR-2000; 2000EP-0105920.
XX
XX (FARB ) BAYER AG.
XX
XX Schmitz G, Bodzioch M;
XX
XX WPI: 2001-640388/74.
XX P-PSDB; AAE13022.
XX
XX New adenosine triphosphate binding cassette transporter-1 gene
XX polymorphisms, useful for diagnosing and treating lipid disorders,
XX cardiovascular diseases and inflammatory diseases
XX
XX Example 1; Fig 1; 48pp; English.
XX
XX The invention relates to four common polymorphisms in the gene encoding
XX ATP-binding cassette transporter-1 (ABCI). ABC1 is associated with
XX decreased ApoA-1 mediated efflux of cholesterol. The polymorphisms in
XX ABC1 directly affects cellular lipid homeostasis, which is a key factor
XX in the atherogenic processes. The ABC1 polymorphisms are useful for
XX diagnosing and treating lipid disorders, cardiovascular diseases
XX (coronary heart disease, atherosclerosis) and inflammatory diseases
XX (psoriasis, lupus erythematosus). The identification of ABC1 as a
XX transporter for interleukin-beta (IL-1beta) identifies this gene as
XX a candidate for treatment of inflammatory diseases including rheumatoid
XX arthritis and septic shock. The present sequence is human ABC1 gene.
XX
XX Sequence 7260 bp; 1834 A; 1765 C; 1905 G; 1756 I; 0 other;
XX
XX Query Match 21.2%; Score 98; DB 22; Length 7260;
XX Best Local Similarity 100.0%; Pred. No. 2.5e-15;
XX Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 366 AAACCCGCTAATTCGACGAGAGTTCAGTGGCGCGGACCGACGAGCGGAGCCGACC 425
XX |||||||
XX Do 1 AAACCCGCTAATTCGACGAGAGTTCAGTGGCGCGGACCGACCGAGCGGAGCCGACC 60
XX |||||||
XX QY 426 TTCTCTCCGGGCTCGGGCAGGCGAGGGCGGGAGCTC 463
XX |||||||
XX Db 61 TTCTCTCCGGGCTCGGGCAGGCGAGGGCGGGAGCTC 98
XX |||||||
XX
RESULT 9
AAI70315
ID AAI70315 standard; cDNA; 7260 BP.
XX

```

```

AC AAI70315;
XX
XX 07-JAN-2002 (first entry)
XX
DE Human ATP binding cassette transporter 1 (ABCI) cDNA.
XX
XX ATP binding cassette transporter 1; ABC1; human; lipid disorder;
KW cholesterol; cardiovascular disease; inflammatory disease;
KW antiinflammatory; antilipemic; arthropathic; dermatological;
KW tangier disease; coronary heart disease; diagnosis; gene therapy;
KW polymorphism; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FI CDS 321..7106
FT CDS /*tag= a
FT CDS 501..7106 /*tag= b
FT FT /*note= "alternative open reading frame of AAI70314"
FT FT /*tag= c
FT variation replace(1516,C)
FT variation /*tag= d
FT variation replace(2969,G)
FT variation /*tag= e
FT variation replace(3836,C)
FT variation /*tag= f
XX
XX EPI136554-A1.
XX
XX 26-SEP-2001.
XX
XX 24-MAR-2000; 2000EP-0106401.
XX
XX 24-MAR-2000; 2000EP-0106401.
XX (FARB ) BAYER AG.
XX
XX Schmitz G, Bodzioch M;
XX
XX WPI: 2001-640389/74.
XX P-PSDB; AAM50228.
XX
XX New adenosine triphosphate binding cassette transporter gene
XX polymorphisms, useful for diagnosing and treating lipid disorders,
XX cardiovascular diseases and inflammatory diseases
XX
XX Disclosure: Page 26-28; 41pp; English.
XX
XX The present sequence is that of cDNA encoding the human adenosine
XX triphosphate (ATP) binding cassette transporter 1 (ABCI) protein
XX (see AAM50227). The sequence includes an extended open reading
XX frame (ORF) to that provided by the sequence in AAI70314, using
XX an alternative ATG codon as initiation codon and thereby adding an
XX extra 40 N-terminal amino acids to the encoded ABC1 protein (see
XX AAM50228). The invention provides 4 common polymorphisms in the
XX ABC1 gene. These were identified by sequencing the ABC1 gene in
XX different tangier kindreds. In the variant genes (numbering as in
XX AAI70314) G is changed to A at position 596, T is changed to C at
XX position 1136, A is changed to G at position 2589 or G is changed
XX to C at position 3456, or any combination of these. All of these
XX polymorphisms alter the amino acid sequence of ABC1 and therefore
XX may affect its function. The 2 most common polymorphisms (G596A
XX and A2589G) are both associated with a decreased in vitro ApoA-I
XX mediated efflux of cholesterol from mononuclear phagocytes, a
XX feature typical of tangier disease. 3 of the variants (G596A, of
XX A2589G and G3456C) are significantly increased in a population of
XX men having low high density lipoprotein-cholesterol levels and
XX established coronary heart disease (CHD) relative to CHD-free
XX control subjects. The use of the provided ABC1 polymorphisms for
XX the diagnosis and treatment of lipid disorders, cardiovascular
XX diseases, and inflammatory diseases (e.g. psoriasis, lupus

```

CC erythematodes) is claimed. Modulation of ABC1 transcripts or
CC proteins by antisense or ribozyme technology or RNA decoys is also
CC claimed.
XX
XX
SQ Sequence 7260 BP; 1834 A; 1765 C; 1905 G; 1756 T; 0 other;
Query Match: 21.2%; Score 98; DP 22; Length 7260;
Best Local Similarity 100.0%; Pred. No. 2.5e-15;
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 366 AAACCCCGTAATTCGAGCGAGAGTGTAGTGGCGCGGACCCGAGCGAGCGACCC 425
DB 1 AAACCCCGTAATTCGAGCGAGAGTGTAGTGGCGCGGACCCGAGCGAGCGACCC 60
QY 426 TTCCTCTCCCGGCTCGCGGCGAGGCGAGGCGGGGAGCTC 453
DB 5. TTCCTCTCCCGGCTCGCGGCGAGGCGAGGCGGGGAGCTC 92
RESULT 10
AAD37268
ID AAD37268 standard; DNA; 221 BP.
XX
XX AAD37268;
XX
XX 21-AUG-2002 (first entry)
XX Human ABC1 gene exon 1A.
DE Human; ATP-binding cassette 1; ABC1 gene regulation; atherosclerosis;
KW cholesterol metabolism; hypercholesterolaemia; antisense therapy;
KW exon 1A; ds.
XX
XX Homo sapiens.
OS
XX WO200183746-A2.
PN
XX 08-NOV-2001.
PD
XX 02-MAY-2001; 2001WO-EP05488.
PF
XX 02-MAY-2000; 2000US-201280P.
PR
XX (AVET) AVENTIS PHARMA SA.
PA
XX Rosier-Montus M, Prades C, Lemoine C, Naudin L, Deneffe P;
PI Brewer B, Duverger N, Remaley A, Sanzamarina-Pojc S;
DR WPI; 2002-154404/20.
XX Isolated nucleic acid useful for modifying the ATP-binding cassette 1
XX (ABC1) and screening for candidate modulatory compounds or substances
XX
XX Claim 4; Page 132; 152pp; English.
PS The invention relates a nucleic acid which is capable of regulating the
XX transcription of human ATP-binding cassette 1 (ABC1) gene, which is a
XX causal gene for pathologies linked to a dysfunctioning of cholesterol
XX metabolism, including diseases such as atherosclerosis. Polynucleotides
XX of the invention are used to screen candidate molecules or substances
XX that are capable of modulating the transcription of the ABC1 gene. They
XX are used in antisense therapy. Compositions comprising sequences of the
XX invention are used to treat hypercholesterolaemia and atherosclerosis.
XX The present sequence is human ABC1 gene exon 1A.
XX
SQ Sequence 221 BP; 44 A; 62 C; 73 G; 42 T; 0 other;
Query Match 19.7%; Score 91; DP 24; Length 221;
Best Local Similarity 100.0%; Pred. No. 8.2e-14;
Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 373 GTAATTCGAGCGAGAGTGTAGTGGCGCGGACCCGAGCGAGCGACCCCTTCTC 432

DB 1 GTAATTCGAGCGAGAGTGTAGTGGCGCGGACCCGAGCGAGCGACCCCTTCTC 60
QY 433 CGGCGCTCGCGGCGAGGCGAGGCGGGGAGCTC 463
DB 61 CGGCGCTCGCGGCGAGGCGAGGCGGGGAGCTC 91
RESULT 11
AAH07432
ID AAH07432 standard; cDNA; 736 BP.
XX
XX AAH07432;
XX
XX 26-JUN-2001 (first entry)
DT Human cDNA clone (5'-primer) SEQ ID NO:4267.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX Homo sapiens.
OS
XX EPI074617-A2.
PN
XX 07-FEB-2001.
PD
XX 28-JUL-2000; 2000EP-0116126.
PF
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-018776.
PR 02-MAY-2000; 2000JP-0183757.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES -NST.
PA
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Negai K, Otsuki T;
XX WPI; 2001-318749/34.
DR
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
PS Claim 1; SEQ ID 4267; 2537pp - CD ROM; English.
XX The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 3'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAH24446 to
XX AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
SQ Sequence 736 BP; 163 A; 199 C; 199 G; 170 T; 5 other;

Query Match 19.3%; Score 89.4; DB 22; Length 736;
Best Local Similarity 98.9%; Pred. No. 2.6e-13;
Matches 90; Conservative 0; Mismatches 1; Indels C; Gaps 0;

XX
SQ Sequence 1556 BP; 380 A; 363 C; 399 G; 414 T; 0 other:

Query Match 19.3%; Score 89.4; DB 22; Length 1556;
Best Local Similarity 98.9%; Pred. No. 2.9e-13;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 373 GTAATTGCGAGCAGGTGGTGGCCGCCGACGCAGAGCCGACCCTTCCTC 432
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 GTAGTTCCGAGCAGTAGTGAGTGGGCCCGGACCCCAGAGCCGACCCCTTCCTC 60

OY 433 CGCGCTGCGGAGGCGAGGCGGGGAGCTC 463
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 CCGGCTGCGGAGGCGAGGCGGGGAGCTC 91

RESULT 12
AAH18606
ID AAH18606 standard; cDNA; 1556 BP.
XX
AC AAH18606;
XX
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:18808.
XX
KW Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX XX
PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
PF
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama I, Wakamatsu A, Nagai K, Otsuki T;
XX WPI: 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

PS Claim 8; SEQ ID 18808; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH19742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 763 BP; 137 A; 205 C; 250 G; 158 T; 3 other;

Query Match 19.2%; Score 89; DB 22; Length 763;
Best Local Similarity 100.0%; Pred. No. 3.3e-13;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 AATTCGAGCGAGAGTGGGCGCGGACCGGAGCGGAGCGGAGCTC 434
|||||
DB 1 AATTCGAGCGAGAGTGGGCGCGGACCGGAGCGGAGCGGAGCTC 60
QY 435 GGGCTGGCGCAGCGGAGCGGAGCGGAGCTC 463
|||||
DB 61 GGGCTGGCGCAGCGGAGCGGAGCGGAGCTC 89

RESULT 14
AAH17451
ID AAH17451 standard; cDNA; 1750 BP.

XX AC AAH-7451;

XX DT 26-JUN-2001 (first entry)

XX DE Human cDNA sequence SEQ ID NO:16905.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX OS Homo sapiens.

XX PN EP1074517-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-0116126.

XX PR 29-JUL-1999; 99JP-0248036.

XX PR 27-AUG-1999; 99JP-0300253.

XX PR 11-JAN-2000; 2000JP-0118776.

XX PR 02-MAY-2000; 2000JP-0183767.

XX PR 09-JUN-2000; 2000JP-0241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama I, Wakamatsu A, Nagai K, Otsuki T;

XX XX WPI; 2001-318749/34.

XX XX Primer sets for synthesizing polynucleotides, particularly the 5602

XX full-length cDNAs defined in the specification, and for the detection

XX and/or diagnosis of the abnormality of the proteins encoded by the

XX full-length cDNAs -

XX Claim 8; SEQ ID 16905; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602

XX full-length cDNAs defined in the specification. Where a primer set

XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesising polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to

CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX Sequence 1750 BP; 291 A; 489 C; 586 G; 384 T; 0 other;

Query Match 19.2%; Score 89; DB 22; Length 1750;
Best Local Similarity 100.0%; Pred. No. 3.3e-13;
Matches 89; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 AATTCGAGCGAGAGTGGGCGCGGACCGGAGCGGAGCGGAGCTC 434
|||||
DB 1 AATTCGAGCGAGAGTGGGCGCGGACCGGAGCGGAGCGGAGCTC 60

QY 435 GGGCTGGCGCAGCGGAGCGGAGCGGAGCTC 463
|||||
DB 61 GGGCTGGCGCAGCGGAGCGGAGCGGAGCTC 89

RESULT 15
AAK51683
ID AAK51683 standard; cDNA; 7281 BP.

XX AC AAK51683;

XX DT 06-NOV-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 225.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;

XX KW nervous system disorder; arthritis; inflammation; ss.

XX OS Homo sapiens.

XX PN WO200157190-A2.

XX PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04092.

XX PR 03-FEB-2000; 2000US-0496914.

XX PR 27-APR-2000; 2000US-0560875.

XX PR 20-JUN-2000; 2000US-0598075.

XX PR 19-JUL-2000; 2000US-0620325.

XX PR 01-SEP-2000; 2000US-0654936.

XX PR 15-SEP-2000; 2000US-0663561.

XX PR 20-OCT-2000; 2000US-0693325.

XX PR 30-NOV-2000; 2000US-0728422.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

XX PI Zhao QA, Wang D, Zhang J, Zhang J, Ren F, Chen R, Wang ZW;

XX PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX XX WPI; 2001-476283/51.

XX DR P-PSDB; AAK78550.

XX Nucleic acids encoding polypeptides with cytokine-like activities,

XX useful in diagnosis and gene therapy -

XX Claim 1; Page 1086-1096; 6221pp; English.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 3, 2003, 16:38:05 : Search time 1007.79 Seconds
(without alignments)
10155.006 Million cell updates/sec

Title: US-09-596-141c-3_COPY_1292_1643

Perfect score: 352

Sequence: 1 ggacctagacaccctctg.....caggccagggcggggagctc 352

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Gen2mb1.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pla.*

35: em_htg_rod.*

36: em_htg_mam.*

37: em_htg_vit.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_hlgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	352	100.0	1167	9	HS252201 Homo sapi
2	352	100.0	1643	6	AX060715 Sequence
3	352	100.0	1643	6	AX060894 Sequence
4	352	100.0	96717	9	AL359182 Human DNA
5	352	100.0	175064	2	AC012230 Homo sapi
6	348.8	99.1	1167	9	AF238623sl
7	339.4	96.4	69570	2	AC021246 Homo sapi
8	336	95.5	183999	6	AX092589 Sequence
9	335.4	95.3	3231	6	AX351029 Sequence
10	335.4	95.3	149034	9	AF275948 Homo sapi
11	333.8	94.8	201144	9	AF287262 Homo sapi
12	271.6	77.2	69570	2	AC021246 Homo sapi
13	244.4	69.4	2893	6	AX351031 Sequence
14	227	64.5	90698	2	AC021345 Homo sapi
15	173	49.1	697	9	AF258627 Homo sapi
16	155.4	44.1	186889	2	AL807243 Mus muscu
17	155.4	44.1	278572	10	AF287263
18	153.2	43.5	145833	2	AC125637 Rattus no
19	98	27.8	7260	6	AX253452 Sequence
20	91	25.9	221	6	AX351032 Sequence
21	89.4	25.4	1556	9	AK024328 Homo sapi
22	89	25.3	1750	9	AX022254 Homo sapi
23	75	21.3	9854	6	AX127831 Sequence
24	75	21.3	9854	6	AX139818 Sequence
25	67	19.0	10442	6	AX060713 Sequence
26	67	19.0	10442	6	AX060892 Sequence
27	67	19.0	10442	9	AF285167 Homo sapi
28	60	17.0	10474	6	AX060719 Sequence
29	60	17.0	10474	6	AX060721 Sequence
30	60	17.0	10474	6	AX060898 Sequence
31	60	17.0	10474	6	AX060900 Sequence
32	43.6	12.4	170792	2	AC112455 Rattus no
33	43.2	12.3	188069	2	AC099251 Rattus no
34	41.4	11.8	136799	2	AC110334 Rattus no
35	40.6	11.5	3867	1	AY033407 Myxococ
36	40.4	11.5	183022	2	AC109103 Rattus no
37	40.2	11.4	182996	2	AC027682 Homo sapi
38	39.6	11.2	125020	9	AF429315 Homo sapi
39	39.2	11.1	162811	2	AC095897 Rattus no
40	39	11.1	207959	2	AC129157 Rattus no
41	38.8	11.0	138151	2	AC094070 Rattus no
42	38.8	11.0	150994	2	AC111413 Rattus no
43	38.6	11.0	146606	2	AC111683 Rattus no
44	38.4	10.9	143330	2	AC094396 Rattus no
45	38.4	10.9	180731	2	AC103182 Rattus no

ALIGNMENTS

RESULT 1
HSA252201 HSA252201 1167 bp DNA linear PRI 10-APR-2001
LOCUS Homo sapiens partial ABC-1 gene for ATP-binding cassette transporter-1, 5'UTR and promoter region.
DEFINITION
ACCESSION AJ252201
VERSION AJ252201.1 GI:12053757
KEYWORDS ABC-1 gene; ATP-binding cassette transporter-1; promoter.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1167)
Porsch-Ozcurumet, M., Langmann, T., Heimerl, S., Borsukova, H., Kaminski, W.E., Drobniak, W., Honer, C., Schumacher, C. and Schmitz, G.

Pred. No. is the number of results predicted by chance to have a

TITLE The zinc finger protein 202 (ZNF202) is a transcriptional repressor of ATP binding cassette transporter A1 (ABCA1) and ABCG1 gene expression and a modulator of cellular lipid efflux

JOURNAL J. Biol. Chem. 276 (15), 12427-12433 (2001)

MEDLINE 21192304

PUBMED 11279331

REFERENCE 2 (bases 1 to 1167)

AUTHORS Porsch-Oezueruemez M.K., Institute for Direct Submission

TITLE Submitted (03-JAN-2000)

JOURNAL Clinical Chemistry, University of Regensburg, Franz-Josef-Strauss-Allee 11, 93042 Regensburg, GERMANY

FEATURES

source

1..1167

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="9"

/map="9q22-31"

/cell_type="leukocyte"

gene 1..1167

/gene="ABC-1"

promoter 1..942

/gene="ABC-1"

/function="cholesterol efflux regulatory protein"

TATA_signal 896..900

/gene="ABC-1"

intron 1148..1167

/gene="ABC-1"

/function="cholesterol efflux regulatory protein"

BASE COUNT 278 a 315 c 327 g 247 t

ORIGIN

Query Match 100.0%; Score 352; DB 9; Length 1167;

Best Local Similarity 100.0%; Pred. No. 2.5e-78;

Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGACCTAGACACCTGCTGTACCTTCACCCACCCACCCACCCACCTCCCTCCCACTCC 60

|||||

Db 666 GGACCTAGACACCTGCTGTACCTTCACCCACCCACCCACCCACCTCCCTCCCACTCC 60

QY 61 TAGATGTCGTGGGGCTGAACCTGCGCCGTTTAAGGGGGGGGGGGGGGGGGTCCACCTG 120

|||||

Db 726 TAGATGTCGTGGGGCTGAACCTGCGCCGTTTAAGGGGGGGGGGGGGGGGGTCCACCTG 120

QY 121 CTTCTGCTGAGTGAACACATACAGAGCGCGGGAACGGGGGGGGGGAGGAGGA 180

|||||

Db 786 CTTCTGCTGAGTGAACACATACAGAGCGCGGGAACGGGGGGGGGGAGGAGGA 180

QY 181 GAGCAGAGCTTTGACCGATAGTAACTCTGCGCTGGTGCAGCGCAATCTATAAAGGA 240

|||||

Db 846 GAGCAGAGCTTTGACCGATAGTAACTCTGCGCTGGTGCAGCGCAATCTATAAAGGA 240

QY 241 ACTAGTCCCGGCAAAACCCCTTAATTGGAGGAGAGTGAAGTGGGGGGGGGGGGGGG 300

|||||

Db 906 ACTAGTCCCGGCAAAACCCCTTAATTGGAGGAGAGTGAAGTGGGGGGGGGGGGGGG 300

QY 301 AGCCGAGCGACCCCTCTCTCCCGGCTCGGCGAGGGGAGGGGGGAGCTC 352

|||||

Db 966 AGCCGAGCGACCCCTCTCTCCCGGCTCGGCGAGGGGAGGGGGAGCTC 1017

RESULT 2

AX060715

LOCUS AX060715 1643 bp DNA linear PAT 22-JAN-2001

DEFINITION Sequence 3 from Patent WO0078971.

ACCESSION AX060715

VERSION AX060715.1 GI:12406204

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM

REFERENCE 1 (bases 1 to 1643)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Lawn,R.M., Wade,D., Oram,J.P. and Garvin,M.

TITLE ATP binding cassette transporter protein abc1 polypeptides

JOURNAL Patent: WO 0078971-A 3 28-DEC-2000;

CV THERAPEUTICS, INC. (US)

FEATURES

source

1..1643

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 370 a 413 c 457 g 403 t

ORIGIN

Query Match 100.0%; Score 352; DB 6; Length 1643;

Best Local Similarity 100.0%; Pred. No. 2.4e-78;

Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGACCTAGACACCTGCTGTACCTTCACCCACCCACCCACCCACCTCCCTCCCACTCC 60

|||||

Db 1292 GGACCTAGACACCTGCTGTACCTTCACCCACCCACCCACCCACCTCCCTCCCACTCC 1351

QY 61 TAGATGTCGTGGGGCTGAACCTGCGCCGTTTAAGGGGGGGGGGGGGGGGGTCCACCTG 120

|||||

Db 1352 TAGATGTCGTGGGGCTGAACCTGCGCCGTTTAAGGGGGGGGGGGGGGGGGTCCACCTG 1411

QY 121 CTTCTGCTGAGTGAACACATACAGAGCGCGGGAACGGGGGGGGGGAGGAGGA 180

|||||

Db 1412 CTTCTGCTGAGTGAACACATACAGAGCGCGGGAACGGGGGGGGGGAGGAGGA 1471

QY 181 GAGCAGAGCTTTGACCGATAGTAACTCTGCGCTGGTGCAGCGCAATCTATAAAGGA 240

|||||

Db 1472 GAGCAGAGCTTTGACCGATAGTAACTCTGCGCTGGTGCAGCGCAATCTATAAAGGA 1531

QY 241 ACTAGTCCCGGCAAAACCCCTTAATTGGAGGAGAGTGAAGTGGGGGGGGGGGGGGG 300

|||||

Db 1532 ACTAGTCCCGGCAAAACCCCTTAATTGGAGGAGAGTGAAGTGGGGGGGGGGGGGGG 300

QY 301 AGCCGAGCGACCCCTCTCTCCCGGCTCGGCGAGGGGAGGGGGGAGCTC 352

|||||

Db 1592 AGCCGAGCGACCCCTCTCTCCCGGCTCGGCGAGGGGAGGGGGAGCTC 1643

AUTHORS Lawn,R.M., Wade,D. and Garvin,M.

TITLE Regulation with binding cassette transporter protein abc1

JOURNAL Patent: WO 0078972-A 3 28-DEC-2000;

CV THERAPEUTICS, INC. (US)

FEATURES

source

1..1643

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 370 a 413 c 457 g 403 t

ORIGIN

Query Match 100.0%; Score 352; DB 6; Length 1643;

Best Local Similarity 100.0%; Pred. No. 2.4e-78;

Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGACCTAGACACCTGCTGTACCTTCACCCACCCACCCACCCACCTCCCTCCCACTCC 60

|||||

Db 1292 GGACCTAGACACCTGCTGTACCTTCACCCACCCACCCACCCACCTCCCTCCCACTCC 1351

QY 61 TAGATGTCGTGGGGCTGAACCTGCGCCGTTTAAGGGGGGGGGGGGGGGGGTCCACCTG 120

|||||

Db 1352 TAGATGTCGTGGGGCTGAACCTGCGCCGTTTAAGGGGGGGGGGGGGGGGGTCCACCTG 1411

QY 121 CTTCTGCTGAGTGAACACATACAGAGCGCGGGAACGGGGGGGGGGAGGAGGA 180

|||||

Db 1412 CTTCTGCTGAGTGAACACATACAGAGCGCGGGAACGGGGGGGGGGAGGAGGA 1471

QY 181 GAGCAGAGCTTTGACCGATAGTAACTCTGCGCTGGTGCAGCGCAATCTATAAAGGA 240

|||||

Db 1472 GAGCAGAGCTTTGACCGATAGTAACTCTGCGCTGGTGCAGCGCAATCTATAAAGGA 1531

QY 241 ACTAGTCCCGGCAAAACCCCTTAATTGGAGGAGAGTGAAGTGGGGGGGGGGGGGGG 300

|||||

Db 1532 ACTAGTCCCGGCAAAACCCCTTAATTGGAGGAGAGTGAAGTGGGGGGGGGGGGGGG 300

QY 301 AGCCGAGCGACCCCTCTCTCCCGGCTCGGCGAGGGGAGGGGGGAGCTC 352

|||||

Db 1592 AGCCGAGCGACCCCTCTCTCCCGGCTCGGCGAGGGGAGGGGGAGCTC 1643

RESULT 3

AX060894

LOCUS AX060894 1643 bp DNA linear PAT 22-JAN-2001

DEFINITION Sequence 3 from Patent WO0078971.

ACCESSION AX060894

VERSION AX060894.1 GI:12406271

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM

REFERENCE 1 (bases 1 to 1643)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Lawn,R.M., Wade,D., Oram,J.P. and Garvin,M.

TITLE ATP binding cassette transporter protein abc1 polypeptides

JOURNAL Patent: WO 0078971-A 3 28-DEC-2000;

CV THERAPEUTICS, INC. (US)

FEATURES

source

1..1643

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 370 a 413 c 457 g 403 t

ORIGIN

Query Match 100.0%; Score 352; DB 6; Length 1643;

Best Local Similarity 100.0%; Pred. No. 2.4e-78;

Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGACCTAGACACCTGCTGTACCTTCACCCACCCACCCACCCACCTCCCTCCCACTCC 60

|||||

Db 1292 GGACCTAGACACCTGCTGTACCTTCACCCACCCACCCACCCACCTCCCTCCCACTCC 1351

QY 61 TAGATGTCGTGGGGCTGAACCTGCGCCGTTTAAGGGGGGGGGGGGGGGGGTCCACCTG 120

|||||

Db 1352 TAGATGTCCTGCGGCGCTGACGCTCCCGCTTATAGGCGGGCGGCCCGGCTCACGTG 1411
 Qy 121 CTTCCTGCTGAGTCTGACTGAATACATAACAGAGCCCGGACCGGGCGGGAGAGGGA 180
 |||||
 Db 1412 CTTCCTGCTGAGTCTGACTGAATACATAACAGAGCCCGGACCGGGCGGGAGAGGGA 1471
 Qy 181 GAGCAGAGGCTTACCGATAGTACCTCTCGCTCCGTCGAGCGGAATCTATAAAGGA 240
 |||||
 Db 1472 GAGCAGAGGCTTACCGATAGTACCTCTCGCTCCGTCGAGCGGAATCTATAAAGGA 1531
 Qy 241 ACTAGTCCCGGCAAAACCCCGTAAITCGGAGCGAGAGTGTAGTGGCGGGACCGCGAG 300
 |||||
 Db 1532 ACTAGTCCCGGCAAAACCCCGTAAITCGGAGCGAGAGTGTAGTGGCGGGACCGCGAG 1591
 Qy 301 AGCGAGCGGACCTTCTCTCCCGGCTGCGGACGAGGCGAGGGCGGGGAGCTC 352
 |||||
 Db 1592 AGCGAGCGGACCTTCTCTCCCGGCTGCGGACGAGGCGAGGGCGGGGAGCTC 1543

RESULT 4
 LOCUS ALJ359182/c 96717 bp DNA linear PRI 11-JAN-2002
 DEFINITION Human DNA sequence from clone RP11-217B7 on chromosome 9, comp-ete
 sequence.
 ACCESSION ALJ359182
 VERSION ALJ359182.20 GI:18151453
 KEYWORDS HTG.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 96717)

StuCe.C
 Direct Submission
 Submitted (11-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Jan 15, 2002 this sequence version replaced gi:18121468.

During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em, EMBL; Sw, SW-SSRPROT; Tr, TRENBL; Wp, WORMPEP; Information on the WORMPEP
 database can be found at

<http://www.sanger.ac.uk/projects/c-elegans/wormpep/> This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr9/>
 RP11-217B7 is from the library RP11-11.1 constructed by the group
 of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>

VECTOR: pBAC3.6
 IMPORTANT: This sequence is not the entire insert of clone
 RP11-217B7. It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true right end of clone RP11-217B7 is at 96717 in this
 sequence. The true left end of clone RP11-22F10 is at 72980 in
 this sequence. The true right end of clone RP11-31J20 is at 2000 in
 this sequence.

FEATURES
 Location/Qualifiers
 1..96717

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /clone="RP11-217B7"
 /clone_id="RP11-11.1"
 3238..3278
 /note="Sequence from AF275948 sequenced by National
 Institutes of Health, National Heart, Lung and Blood
 Institute, Bethesda, MD 20892, USA."
 84249..84273
 /note="Sequence from overlapping clone RP11-122F10
 (AC026643). Assembly confirmed by restriction digest."
 92050..92163
 /note="Sequence from reads from a short insert library
 derived from a single pUC clone. Restriction digest data
 confirm the assembly."
 92411..92557
 /note="Sequence from reads from a short insert library
 derived from a single pUC clone. Restriction digest data
 confirm the assembly."
 27673 a 21138 c 20380 g 27526 t

BASE COUNT 27673 a 21138 c 20380 g 27526 t
 ORIGIN
 Query Match 100.0%; Score 352; DB 9; Length 96717;
 Best Local Similarity 100.0%; Pred. No. 1,6e-78;
 Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGACCTAAGACACCTGTGTACCTCCACCCACCCACCCACCTCCCTCCCACTCC 60
 |||||
 Db 17518 GGACCTAAGACACCTGTGTACCTCCACCCACCCACCCACCTCCCTCCCACTCC 17459
 Qy 61 TAGATGTCGCTGGCGGCTGACCTGCGCCGCTTAAAGGGCGGGCGGCTCCACCTG 120
 |||||
 Db 17458 TAGATGTCGCTGGCGGCTGACCTGCGCCGCTTAAAGGGCGGGCGGCTCCACCTG 17399
 Qy 121 CTTCTGCTGAGTCTGAACCTACATAACAGAGCCCGGAGCGGGGAGGAGGGA 180
 |||||
 Db 17398 CTTCTGCTGAGTCTGAACCTACATAACAGAGCCCGGAGCGGGGAGGAGGGA 17339
 Qy 181 GAGCAGAGCTTTGACCCATAGTAACTCTGCGCTGCTGAGCGGAATCTATAAAGGA 240
 |||||
 Db 17338 GAGCAGAGCTTTGACCCATAGTAACTCTGCGCTGCTGAGCGGAATCTATAAAGGA 17278
 Qy 241 ACTAGTCCCGGCAAAACCCCGTAAITCGGAGCGAGAGTGTAGTGGCGGGACCGCGAG 300
 |||||
 Db 17278 ACTAGTCCCGGCAAAACCCCGTAAITCGGAGCGAGAGTGTAGTGGCGGGACCGCGAG 17219
 Qy 301 AGCGAGCGGACCTTCTCTCCCGGCTGCGGACGAGGCGAGGGCGGGGAGCTC 352
 |||||
 Db 17218 AGCGAGCGGACCTTCTCTCCCGGCTGCGGACGAGGCGAGGGCGGGGAGCTC 17167

RESULT 5
 AC012230
 LOCUS

DEFINITION Homo sapiens clone RP11-1M10, WORKING DRAFT SEQUENCE, 39 unordered
 pieces.
 ACCESSION AC012230
 VERSION AC012230.3 GI:7637254
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 175064)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 TITLE Homo sapiens, clone RP11-1M10
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 175064)
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
 Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
 Cooke,P., Dearellano,K., Dewar,K., Domino,K., Donelan,L., Doyle,M.,

Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Hcrlton, L.,
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McEwan, K., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stefanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.

TITLE JOURNAL

Submitted (21-OCT-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 22, 2000 this sequence version replaced gi:6454033.

COMMENT

All repeats were identified using RepeatMasker:
 Smit, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L251C

Center clone name: L_M10

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Parap; version 0.960731

Consensus quality: 117571 bases at least Q40

Consensus quality: 145749 bases at least Q30

Consensus quality: 160940 bases at least Q20

Insert size: 185000; agarose-fo

Insert size: 171264; sum-of-contigs

Quality coverage: 2.9 in Q20 bases; agarose-fo

Quality coverage: 3.2 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 39 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 1003: contig of 1003 bp in length
1004 1103: gap of 100 bp
1104 2634: contig of 1531 bp in length
2635 2734: gap of 100 bp
2735 4415: contig of 1681 bp in length
4416 4515: gap of 100 bp
4516 5785: contig of 1270 bp in length
5786 5885: gap of 100 bp
5886 7879: contig of 1994 bp in length
7880 7979: gap of 100 bp
7980 9686: contig of 1707 bp in length
9687 9786: gap of 100 bp
9787 12253: contig of 2467 bp in length
12254 12353: gap of 100 bp
12354 15228: contig of 2875 bp in length
15229 15287: gap of 100 bp
15288 17200: contig of 1972 bp in length
17201 17300: gap of 100 bp
17301 20131: contig of 2831 bp in length
20132 20231: gap of 100 bp
20232 22587: contig of 2356 bp in length
22588 22687: gap of 100 bp
22688 25707: contig of 3020 bp in length
25708 25807: gap of 100 bp
25808 28184: contig of 2377 bp in length
28185 28284: gap of 100 bp
28285 31338: contig of 3054 bp in length
31339 32438: gap of 100 bp

```

```

* 31439 34299: contig of 2861 bp in length
* 34300 34399: gap of 100 bp
* 34400 3818: contig of 3919 bp in length
* 3819 38418: gap of 100 bp
* 38419 42835: contig of 4417 bp in length
* 42836 42935: gap of 100 bp
* 42936 45448: contig of 2533 bp in length
* 45449 45548: gap of 100 bp
* 45549 48116: contig of 2568 bp in length
* 48117 48216: gap of 100 bp
* 48217 52718: contig of 4402 bp in length
* 52719 56592: contig of 3874 bp in length
* 56593 56692: gap of 100 bp
* 56693 59635: contig of 2943 bp in length
* 59636 59735: gap of 100 bp
* 59736 63661: contig of 3926 bp in length
* 63662 63761: gap of 100 bp
* 63762 68437: contig of 4676 bp in length
* 68438 68537: gap of 100 bp
* 68538 71458: contig of 2921 bp in length
* 71459 71558: gap of 100 bp
* 71559 76888: contig of 5330 bp in length
* 76889 76988: gap of 100 bp
* 76989 82113: contig of 5125 bp in length
* 82114 82213: gap of 100 bp
* 82214 88220: contig of 6007 bp in length
* 88221 89320: gap of 100 bp
* 89321 93499: contig of 5179 bp in length
* 93500 93599: gap of 100 bp
* 93600 97901: contig of 4302 bp in length
* 97902 98001: gap of 100 bp
* 98002 103016: contig of 5015 bp in length
* 103017 103116: gap of 100 bp
* 103117 109798: contig of 6662 bp in length
* 109799 109278: gap of 100 bp
* 109279 117307: contig of 8029 bp in length
* 117308 117407: gap of 100 bp
* 117408 124079: contig of 6672 bp in length
* 124080 124179: gap of 100 bp
* 124180 132281: contig of 7102 bp in length
* 132282 131381: gap of 100 bp
* 131382 138059: contig of 6678 bp in length
* 138060 138159: gap of 100 bp
* 138160 145491: contig of 7332 bp in length
* 145492 145591: gap of 100 bp
* 145592 157391: contig of 11800 bp in length
* 157392 157491: gap of 100 bp
* 157492 175064: contig of 17573 bp in length.

```

FEATURES

Source

```

1. 175064
/organism="Homo sapiens"
/doi="taxon:9606"
/clone="RP11-1M10"
/clone_lib="RPC1-11 Human Male BAC"
1. 11003

```

misc_feature

```

1104..2634
/note="assembly_fragment"

```

misc_feature

```

2735..4415
/note="assembly_fragment"

```

misc_feature

```

4516..5785
/note="assembly_fragment"

```

misc_feature

```

5886..7879
/note="assembly_fragment"

```

misc_feature

```

7980..9686
/note="assembly_fragment"

```

misc_feature

```

9787..12253
/note="assembly_fragment"

```

misc_feature

```

12354..15228
/note="assembly_fragment"

```

misc_feature

```

15329..17200
/note="assembly_fragment"

```

```

misc_feature 17301..20131
             /note="assembly_fragment"
misc_feature 20232..22587
             /note="assembly_fragment"
misc_feature 22688..25707
             /note="assembly_fragment"
misc_feature 25808..28184
             /note="assembly_fragment"
misc_feature 28285..31338
             /note="assembly_fragment"
misc_feature 31439..34299
             /note="assembly_fragment"
misc_feature 34400..38318
             /note="assembly_fragment"
misc_feature 38419..42835
             /note="assembly_fragment"
misc_feature 42936..45448
             /note="assembly_fragment"
misc_feature 45549..48116
             /note="assembly_fragment"
misc_feature 48217..52618
             /note="assembly_fragment"
misc_feature 52719..56592
             /note="assembly_fragment"
misc_feature 56693..59635
             /note="assembly_fragment"
misc_feature 59736..63661
             /note="assembly_fragment"
misc_feature 63762..68437
             /note="assembly_fragment"
misc_feature 68538..71458
             /note="assembly_fragment"
misc_feature 71559..76888
             /note="assembly_fragment"

Query Match      100.0%; Score 352; DB 2; Length 175064;
Best Local Similarity 100.0%; Pred. No. 1.5e-78;
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGACCTTAAGACACCTGCTGTACCTCCACCCACCCACCCACCCACCTCCCTCCCACTCCC 60
Db 3401 GGACCTTAAGACACCTGCTGTACCTCCACCCACCCACCCACCCACCTCCCTCCCACTCCC 3460

Qy 61 TAGATGCTGCTGGCGGCTGAACCTCCGCGTTTAAAGGGGGGGGGGGGGGGGGGGGGGG 120
Db 3461 TAGATGCTGCTGGCGGCTGAACCTCCGCGTTTAAAGGGGGGGGGGGGGGGGGGGGGGG 3520

Qy 121 CTTTCTGCTGAGTGAACCTACATAACAGAGCCGCGGGAACGGGGGGGGGAGGAGGA 180
Db 3521 CTTTCTGCTGAGTGAACCTACATAACAGAGCCGCGGGAACGGGGGGGGGAGGAGGA 3580

Qy 181 GAGCACAGGCTTTGACCGATAGTAACCTTCGCGTTCGAGCCGGAATCTATAAAAGA 240
Db 3581 GAGCACAGGCTTTGACCGATAGTAACCTTCGCGTTCGAGCCGGAATCTATAAAAGA 3640

Qy 241 ACTAGTCCCGGCAAAACCCCGTATTCGGACGAGAGTGAAGTGGGGGGGGGAGCCGAG 300
Db 3641 ACTAGTCCCGGCAAAACCCCGTATTCGGACGAGAGTGAAGTGGGGGGGGGAGCCGAG 3700

Qy 301 AGCCGAGCGGACCCCTCTCTCCCGGGCTCGCGCAGGGGAGGGGGGAGCTC 352
Db 3701 AGCCGAGCGGACCCCTCTCTCCCGGGCTCGCGCAGGGGAGGGGGGAGCTC 3752

RESULT 6
LOCUS AF258623S1 1167 bp DNA linear PRI 23-JUN-2000
DEFINITION Homo sapiens ATP binding cassette transporter - (ABCA1) gene,
            promoter and exon 1.
ACCESSION AF258623
VERSION AF258623.2 GI:8677405
KEYWORDS

```

```

SEGMENT 1 of 4
SOURCE Homo sapiens.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1167)
AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C.,
          Aouizerat,B.E., Fielding,C.J. and Kane,J.P.
TITLE Analysis of ABC1 gene 5' end: additional peptide sequence,
          promoter region, and four polymorphisms
JOURNAL Biochem. Biophys. Res. Commun. 271 (2000) In press
REFERENCE 2 (bases 224 to 1167)
AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C.,
          Aouizerat,B.E., Fielding,C.J. and Kane,J.P.
TITLE Direct Submission
JOURNAL Submitted (20-APR-2000) Cardiovascular Research Institute,
          University of California, San Francisco, 505 Parnassus Avenue, San
          Francisco, CA 94143-0130, USA
REFERENCE 3 (bases 1 to 1167)
AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C.,
          Aouizerat,B.E., Fielding,C.J. and Kane,J.P.
TITLE Direct Submission
JOURNAL Submitted (23-JUN-2000) Cardiovascular Research Institute,
          University of California, San Francisco, 505 Parnassus Avenue, San
          Francisco, CA 94143-0130, USA
REMARK Sequence update by submitter
COMMENT On Jun 23, 2000 this sequence version replaced gi:7769713.
FEATURES
          Location/Qualifiers
            source
              1..1167
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="9"
                /map="9q31"
                /gene="ABCA1"
                /gene="ABCA1"
                /number=1
            promoter
            exon
            BASE COUNT 278 a 313 c 328 g 248 t
            ORIGIN
              Query Match      99.1%; Score 348.8; DB 9; Length 1167;
              Best Local Similarity 99.4%; Pred. No. 1.6e-77;
              Matches 350; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

              Qy 1 GGACCTTAAGACACCTGCTGTACCTCCACCCACCCACCCACCTCCCTCCCACTCCC 60
              Db 566 GGACCTTAAGACACCTGCTGTACCTCCACCCACCCACCCACCTCCCTCCCACTCCC 725

              Qy 61 TAGATGCTGCTGGCGGCTGAACCTCCGCGTTTAAAGGGGGGGGGGGGGGGGGGGGGGG 120
              Db 726 TAGATGCTGCTGGCGGCTGAACCTCCGCGTTTAAAGGGGGGGGGGGGGGGGGGGGGGG 785

              Qy 121 CTTTCTGCTGAGTGAACCTACATAACAGAGCCGCGGGAACGGGGGGGGGAGGAGGA 180
              Db 786 CTTTCTGCTGAGTGAACCTACATAACAGAGCCGCGGGAACGGGGGGGGGAGGAGGA 845

              Qy 181 GAGCACAGGCTTTGACCGATAGTAACCTTCGCGTTCGAGCCGGAATCTATAAAAGA 240
              Db 846 GAGCACAGGCTTTGACCGATAGTAACCTTCGCGTTCGAGCCGGAATCTATAAAAGA 905

              Qy 241 ACTAGTCCCGGCAAAACCCCGTATTCGGACGAGAGTGAAGTGGGGGGGGGAGCCGAG 300
              Db 906 ACTAGTCCCGGCAAAACCCCGTATTCGGACGAGAGTGAAGTGGGGGGGGGAGCCGAG 965

              Qy 301 AGCCGAGCGGACCCCTCTCTCCCGGGCTCGCGCAGGGGAGGGGGGAGCTC 352
              Db 966 AGCCGAGCGGACCCCTCTCTCCCGGGCTCGCGCAGGGGAGGGGGGAGCTC 1017

RESULT 7
AC021246
LOCUS AC021246 69570 bp DNA linear HTG 13-JUL-2000

```

```

DEFINITION Homo sapiens clone RP11-IN10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC021246
VERSION AC021246.2 GI:9119882
KEYWORDS HTG; HTGS_PHASE0.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 69570)
Barren,B., Linton,L., Nusbaum,C., Allen,N.,
Barren,B., Linton,L., Nusbaum,C., Lande,E., Abraham,H., Allen,N.,
Anderson,S., Badway,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,I., Bokshgalter,B., Brown,A., Burket,G., Castle,A.,
Choel,Y., Collangelo,K., Collins,S., Collymore,A., Cooke,P.,
DeArelano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferrel,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,A., Jones,C., Kan,J., Karatas,A., Klein,J.,
Landers,T., Leoczky,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheters,R., Melidrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Zimmet,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Lander,A. and Zody,M.
Direct Submission
Submitted (15-JAN-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6705871.
All repeats were identified using RepeatMasker:
Sait, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W13R
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2512
Center clone name: 1_N_10
-----
* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 871: contig of 871 bp in length
872 971: gap of 100 bp
972 1834: contig of 863 bp in length
1835 1934: gap of 100 bp
1935 2804: contig of 870 bp in length
2805 2904: gap of 100 bp
2905 3745: contig of 841 bp in length
3745 3845: gap of 100 bp
3845 4696: contig of 851 bp in length
4697 4796: gap of 100 bp
4797 5640: contig of 844 bp in length
5641 5740: gap of 100 bp
5741 6540: contig of 800 bp in length
6541 6640: gap of 100 bp
6641 7509: contig of 869 bp in length
7510 7609: gap of 100 bp
*
* 8710 8473: contig of 970 bp in length
* 8480 8579: gap of 100 bp
* 8580 9430: contig of 851 bp in length
* 9431 9530: gap of 100 bp
* 9531 10376: contig of 846 bp in length
* 10377 10476: gap of 100 bp
* 10477 11322: contig of 846 bp in length
* 11323 11422: gap of 100 bp
* 11423 12302: contig of 830 bp in length
* 12303 12402: gap of 100 bp
* 12403 13280: contig of 878 bp in length
* 13281 13380: gap of 100 bp
* 13381 14241: contig of 861 bp in length
* 14242 14341: gap of 100 bp
* 14342 15196: contig of 855 bp in length
* 15197 15296: gap of 100 bp
* 15297 16123: contig of 827 bp in length
* 16124 16223: gap of 100 bp
* 16224 17072: contig of 849 bp in length
* 17073 17172: gap of 100 bp
* 17173 18041: contig of 869 bp in length
* 18042 18141: gap of 100 bp
* 18142 19009: contig of 868 bp in length
* 19010 19109: gap of 100 bp
* 19110 19966: contig of 857 bp in length
* 19967 20066: gap of 100 bp
* 20067 20921: contig of 855 bp in length
* 20922 21021: gap of 100 bp
* 21022 21866: contig of 844 bp in length
* 21866 21965: gap of 100 bp
* 21966 22832: contig of 867 bp in length
* 22833 22932: gap of 100 bp
* 22933 23780: contig of 848 bp in length
* 23781 23880: gap of 100 bp
* 23881 24733: contig of 853 bp in length
* 24734 24833: gap of 100 bp
* 24834 25670: contig of 837 bp in length
* 25671 25770: gap of 100 bp
* 25771 26621: contig of 851 bp in length
* 26622 26721: gap of 100 bp
* 26722 27576: contig of 855 bp in length
* 27577 27676: gap of 100 bp
* 27677 28532: contig of 856 bp in length
* 28533 28632: gap of 100 bp
* 28633 29492: contig of 860 bp in length
* 29493 29592: gap of 100 bp
* 29593 30455: contig of 863 bp in length
* 30456 30555: gap of 100 bp
* 30556 31410: contig of 855 bp in length
* 31411 31510: gap of 100 bp
* 31511 32368: contig of 858 bp in length
* 32369 32468: gap of 100 bp
* 32469 33312: contig of 844 bp in length
* 33313 33412: gap of 100 bp
* 33413 34268: contig of 856 bp in length
* 34269 34368: gap of 100 bp
* 34369 35204: contig of 836 bp in length
* 35205 35304: gap of 100 bp
* 35305 36156: contig of 852 bp in length
* 36157 36256: gap of 100 bp
* 36257 37128: contig of 872 bp in length
* 37129 37228: gap of 100 bp
* 37229 38083: contig of 855 bp in length
* 38084 38183: gap of 100 bp
* 38184 39031: contig of 848 bp in length
* 39032 39131: gap of 100 bp
* 39132 40006: contig of 875 bp in length
* 40007 40106: gap of 100 bp
* 40107 40967: contig of 861 bp in length
* 40968 41067: gap of 100 bp
* 41068 41913: contig of 846 bp in length
* 41914 42013: gap of 100 bp
* 42014 42824: contig of 811 bp in length

```

```
* 42825 42924: gap of 100 bp
* 42925 43776: contig of 852 bp in length
* 43777 43876: gap of 100 bp
* 43877 44752: contig of 876 bp in length
* 44753 44852: gap of 100 bp
* 44853 44724: contig of 872 bp in length
* 45725 45824: gap of 100 bp
* 45825 45643: contig of 819 bp in length
* 45644 46743: gap of 100 bp
* 46744 47599: contig of 856 bp in length
* 47600 47698: gap of 100 bp
* 47700 48551: contig of 852 bp in length
* 48552 48651: gap of 100 bp
* 48652 49485: contig of 834 bp in length
* 49486 49585: gap of 100 bp
* 49586 50440: contig of 855 bp in length
* 50441 50540: gap of 100 bp
* 50541 51404: contig of 864 bp in length
* 51405 51504: gap of 100 bp
* 51505 52372: contig of 868 bp in length
* 52373 52472: gap of 100 bp
* 52473 53328: contig of 856 bp in length
* 53329 53428: gap of 100 bp
* 53429 54268: contig of 840 bp in length
* 54269 54368: gap of 100 bp
* 54369 55228: contig of 861 bp in length
* 55230 55329: gap of 100 bp
* 55330 55197: contig of 888 bp in length
* 56198 56297: gap of 100 bp
* 56298 57163: contig of 866 bp in length
* 57164 57263: gap of 100 bp
* 57264 58130: contig of 867 bp in length
* 58131 59230: gap of 100 bp
* 59231 59082: contig of 852 bp in length
* 59083 59182: gap of 100 bp
* 59183 60020: contig of 838 bp in length
* 60021 60120: gap of 100 bp
* 60121 60983: contig of 863 bp in length
* 60984 61083: gap of 100 bp
* 61084 61935: contig of 852 bp in length
* 61936 62035: gap of 100 bp
* 62036 62866: contig of 831 bp in length
* 62867 62966: gap of 100 bp
* 62967 63827: contig of 861 bp in length
* 63828 63927: gap of 100 bp
* 63928 64783: contig of 856 bp in length
* 64784 64883: gap of 100 bp
* 64884 65740: contig of 857 bp in length
* 65741 65840: gap of 100 bp
* 65841 66684: contig of 844 bp in length
* 66685 66784: gap of 100 bp
* 66785 67651: contig of 867 bp in length
* 67652 67751: gap of 100 bp

Query Match
Best Local Similarity 96.4%; Score 339.4; DB 2; Length 69570;
Matches 351; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 GGACCTTAAGACACCTGCTGTACCTCCACCCACCCACCCACCTCCCGCCCAACT-CC 59
|||||
DB 41303 GGACCTTAAGACACCTGCTGTACCTCCACCCACCCACCCACCTCCCGCCCAACTCC 41362
|||||

QY 60 CTATATGTCGTCGGGGGGCTGAAGTCGGCGGTTAAGGGGGGGGGCGCGCGCTCCAGT 119
|||||
DB 41363 CTATATGTCGTCGGGGGGCTGAAGTCGGCGGTTAAGGGGGGGGGCGCGCTCCAGT 41422
|||||

QY 120 GCTTCTCTCTGAGTGACGTAACATACATTAACAGAGCGCGGAAAGGGGGGGGAGGAGG 179
|||||
DB 41423 GCTTCTCTCTGAGTGACGTAACATACATTAACAGAGCGCGGAAAGGGGGGGGAGGAGG 41482
|||||

QY 180 AGAGCACAGGCTTTGACCGCAAGTAACCTCTCGGCTCGGTGACGCCGCAATCTATAAAGG 239
|||||
DB 41483 AGAGCACAGGCTTTGACCGCAAGTAACCTCTCGGCTCGGTGACGCCGCAATCTATAAAGG 41542
|||||
```

```
QY 240 AACTAGTCCCGGCAAAACCCCGTAATTCGAGCGAGAGCTGAGTGGGGCGGAGCCGCA 299
|||||
DB 41543 AACTAGTCCCGGCAAAACCCCGTAATTCGAGCGAGAGCTGAGTGGGGCGGAGCCGCA 41602
|||||

QY 300 GAGCGAGCGGACCTTCTCICCGGCGCGGCGGAGGCGGCGGAGGCTC 352
|||||
DB 41603 GAGCGAGCGGACCTTCTCICCGGCGCGGCGGAGGCGGCGGAGGCTC 41655
|||||

RESULT 8
LOCUS AX052589 183999 bp DNA linear PAT 21-MAR-2001
DEFINITION Sequence 1 from Patent WO0115676.
ACCESSION AX052589
VERSION AX052589.1 GI:13444647
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 183999)
AUTHORS Hayden,M.R., Brooks-Wilson,A.R., Pimstone,S.N. and Clee,S.M.
TITLE Compositions and methods for modulating hdl cholesterol and
triglyceride levels
JOURNAL Patent: WO 0115676-A 1 08-MAR-2001;
FEATURES
source
location/Qualifiers
:organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 49549 a 37944 c 41170 g 54950 t
ORIGIN
Query Match 95.5%; Score 336; DB 6; Length 183999;
Best Local Similarity 98.3%; Pred. No. 1.6e-74;
Matches 351; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY 1 GGACCTTAAGACACCTGCTGTACCTCCA-----CCCCACCCACCCACCTCCCGCCCAA 55
|||||
DB 28524 GGACCTTAAGACACCTGCTGTACCTCCANNNNCCACCCACCCACCTCCCGCCCAA 28583
|||||

QY 56 CTCCTAGATGTGCTGGGGGCTGAACGTGCGCGCTTAAGGGGCGGCGCGCGCTCC 115
|||||
DB 28584 CTCCTAGATGTGCTGGGGGCTGAACGTGCGCGCTTAAGGGGCGGCGCGCGCTCC 28643
|||||

QY 116 ACGTCTTTCTGCTGAGTGAAGTAACTACATAAAGAGGCGGGAACGGGCGGGGAGG 175
|||||
DB 28644 ACGTCTTTCTGCTGAGTGAAGTAACTACATAAAGAGGCGGGAACGGGCGGGGAGG 28703
|||||

QY 176 AGGAGACACAGGCTTTGACCGGATAGTAACCTCTGCGCTCGGTGACGCCGAATCTATAA 235
|||||
DB 28704 AGGAGACACAGGCTTTGACCGGATAGTAACCTCTGCGCTCGGTGACGCCGAATCTATAA 28763
|||||

QY 236 AAGGAACCTAGTCCCGGCAAAACCCCGTAATTCGAGCGAGAGCTGAGTGGGGCGGAGCC 295
|||||
DB 28754 AAGGAACCTAGTCCCGGCAAAACCCCGTAATTCGAGCGAGAGCTGAGTGGGGCGGAGCC 28823
|||||

QY 296 CGCAGCGCGAGCGGACCCCTTCTCTCCCGGCTGCGGAGCGGCGGAGGCGGAGCTC 352
|||||
DB 28824 CGCAGCGCGAGCGGACCCCTTCTCTCCCGGCTGCGGAGCGGCGGAGGCGGAGCTC 28880
|||||

RESULT 9
LOCUS AX351029 3231 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 1 from Patent WO0183746.
ACCESSION AX351029
VERSION AX351029.1 GI:18616385
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

REFERENCE		FEATURES		Location/Qualifiers	
AUTHORS	1	Mammalia: Eutheria; Primates; Catarrhini; Hominidae; Homo.	Source	..149034	
		Rosier-Montus M.F., Prades C., Lemoine C., Naudin L., Deneffe P.,		/organism="Homo sapiens"	
		Brewer B., Duverger N., Remaley A. and Santamarina-Fojo S.	Gene	/db_xref="taxon:9606"	
		Patent: WO 0183746-A 1 08-NOV-2001;	mRNA	/gene="ABCA1"	
FEATURES		Aventis Pharma S.A. (FR)		Join(1454..1674,25831..25989,40385..40478,45012..45153,	
		Location/Qualifiers		14623..46541,67729..67850,70831..71007,83960..84052,	
		1..3231		89010..89250,91962..92101,92433..92549,96758..96955,	
		/organism="Homo sapiens"		97702..97907,98428..98604,100391..100613,102360..102581,	
BASE COUNT		809 a 773 c 876 g 773 t		103642..103846,104951..105064,106862..107033,	
				108023..108154,109460..109602,109806..109943,	
				110646..110866,112124..112196,113183..113385,	
				115053..115101,115287..115420,116806..116954,	
ORIGIN				118804..118726,119984..120062,123094..123283,	
				124804..124898,126208..126240,127363..127468,	
				128943..129017,129539..129708,130939..131116,	
				133114..133229,133341..133485,134526..134649,	
Query Match		95.3%; Score 335.4; DB 6; Length 3231;		135736..135865,136130..136250,137038..137100,	
		Best Local Similarity 98.3%; Pred. No. 3.4e-74;		138008..138114,140468..140610,140982..141116,	
		Matches 351; Conservative 0; Mismatches 1; Indels 5; Gaps 1;		142080..142163,142646..142738,143397..143640,	
				144581..148034)	
QY		2 GGACCTAGACACCGTGTACCTCCAC-----CCCCACCCACCCACCTCCGCCCAA 55		/gene="ABCA1"	
				/product="ABCA1"	
		Db 2628 GGACCTAGACACCGTGTACCTCCACCCACCCACCCACCTCCGCCCAA 2687		Join(1454..1674,25831..25923)	
				/gene="ABCA1"	
QY		56 CTCCTAGATGTCTGTCGCGCGCTGAACGTGCGCCGCTTAAGGCGCGCGCGCTCC 115	5'UTR	2406..2678	
			repeat_region	/rpt_family="Alu"	
		Db 2688 CTCCTAGATGTCTGTCGCGCGCTGAACGTGCGCCGCTTAAGGCGCGCGCGCTCC 2747	repeat_region	4240..4509	
			repeat_region	4598..5270	
QY		116 ACCTGCTTCTCTGCTAGTGAAGTGAACATAAAGAGCGCGGAGCGGGGAGG 175	repeat_region	5563..5839	
			repeat_region	/rpt_family="Alu"	
		Db 2748 ACCTGCTTCTCTGCTAGTGAAGTGAACATAAAGAGCGCGGAGCGGGGAGG 2807	repeat_region	6615..6877	
			repeat_region	8800..9046	
QY		176 AGGAGACACAGGCTTACCGATAGTAACTCTCGGTGCGTGCAGCGCAATCTATAA 235	repeat_region	998..10279	
			repeat_region	/rpt_family="Alu"	
		Db 2808 AGGAGACACAGGCTTACCGATAGTAACTCTCGGTGCGTGCAGCGCAATCTATAA 2867	repeat_region	11859..12133	
			repeat_region	12810..12902	
QY		236 AAGGACTAGTCCGCGCAAAACCCCGTAAATGCGAGCGAGAGTGAAGTGGCGCGGACC 295	repeat_region	15220..15399	
			repeat_region	/rpt_family="Alu"	
		Db 2868 AAGGACTAGTCCGCGCAAAACCCCGTAAATGCGAGCGAGAGTGAAGTGGCGCGGACC 2927	repeat_region	15508..15787	
			repeat_region	18600..18860	
QY		296 CCGAGCGCGAGCGGACCTCTCTCCCGGCTGCGGCGAGCGCGGCGGAGCTC 352	repeat_region	20950..21206	
			repeat_region	/rpt_family="Alu"	
		Db 2928 CCGAGCGCGAGCGGACCTCTCTCCCGGCTGCGGCGAGCGCGGCGGAGCTC 2984	repeat_region	25039..25314	
			repeat_region	/rpt_family="Alu"	
RESULT 10		AF275948	CDS	Join(25924..25989,40385..40478,45012..45153,46423..46541,	
		LOCUS		67729..67850,70831..71007,83960..84052,89010..89250,	
		DEFINITION		91962..92101,92433..92549,96758..96955,97702..97907,	
		ACCESSION		98428..98604,100391..100613,102360..102581,103642..103846,	
ACCESSION		AF275948		104351..105064,106862..107033,108023..108154,	
		VERSION		109460..109602,109806..109943,110646..110866,	
		KEYWORDS		112124..112196,113183..113385,115053..115101,	
		SOURCE		115297..115410,116806..116954,118604..118728,	
ORGANISM		Homo sapiens.		119964..120062,123094..123283,124804..124898,	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		126208..126240,127363..127468,128943..129017,	
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		129539..129708,130939..131116,133114..133229,	
		1 (bases 1 to 149034)		13341..133485,134526..134649,135736..135865,	
REFERENCE		Santamarina-Fojo S., Peterson K., Knapper C., Qiu Y., Freeman L.,		136330..136250,137038..137100,138008..138114,	
		Cheng J.F., Osorio J., Remaley A., Yang X.P., Haudenschield C.,		140469..140610,140982..141116,142060..142163,	
		Prades C., Chimini G., Blackmon E., Francois T., Duverger N.,		142646..142738,143397..143640,144581..144721)	
		Rubin E.M., Rosier M., Deneffe P., Fredrickson D.S. and Brewer H.B. Jr.		/gene="ABCA1"	
AUTHORS		Complete genomic sequence of the human ABCA1 gene: analysis of the		/codon_start=1	
		human and mouse ATP-binding cassette A promoter		/product="ABCA1"	
		Proc. Natl. Acad. Sci. U.S.A. 97 (14), 7987-7992 (2000)		/protein_id="AAF66276.1"	
		20345099			
MEDLINE		10864428			
		PUBMED			
		REFERENCE			
		2 (bases 1 to 149034)			
SUBMISSION		Santamarina-Fojo S., Peterson K.M., Knapper C.L., Freeman J.A.,			
		Remaley A.P., Yang X.-P., Haudenschield C.C., Blackmon E.B.,			
		Francois T.L. and Brewer H.B. Jr.			
		Submitted (08-JUN-2000) Molecular Disease Branch, National			
TITLE		Institutes of Health, National Heart, Lung and Blood Institute,			
		Bethesda, MD 20892, USA			

/db_xref="GI:1247086"
 translation="MACPQURLLLXNLTFRRSQTCLLEVAWPLFIILISLVS
 SYPPVEHECHFNPKMPCFSACTLPWQVICNANNPCFRYPTEGAEVGVNGNFKSVL
 ARLSDSARLLLYSQKDTSMKDKVLRIQQIKSSNNKLQDFLQNDVTFSGVLYH
 NLSLPKSTVDKMLRADVILHKVFGQVQLHLLSCNGSKSEMIQVCDQVSELCGLP
 REKLAAEVRLRNNDLKPILRLTISPTPPSKAEAFATKLLHSLGTLAQELFASNR
 SWSMRQCEMFLTVNWSSSSTOIQVRSRVCQEPGGGLKKSLLWYEDNKKALF
 GGNTEDEAEFTYDNSTIPIYCNLMKNLESPLSRVCOGEPGGGLKKSLLWYEDNKKALF
 QVMAVNNKTQELAAFHOLEGMEDESPKWTFMENSOEOMVRLMDSRMDHWE
 QOLGDLWTAQDIWAFLAKHPEECCSGLSYTWEAFNETNQAIRISRFMCVNLN
 KLEPIATFVWLINKSMELLDRKKWAGVIYTGIPGSGILEPHVKKYIRMDINVERT
 NKIKDGYWPGPRADPEEDMYVWGGAIVQIDVVEQALIRLTCTEKKTKGVYQWQPY
 PWCYVDIIFLYMSRGLMFTLAMIYVANKIKGIVYKPEARKETMRKIMGLNSILM
 PSWYTSILPILYVSGLLVILKGNLLPYSDPSVVFVLSVFAVVIITLOCLITSLF
 SRANLACAGGIIITFLIPIVLVAMODIVGFTLKFAXLLSPVAFSGCEVFALPE
 YQGVQVMDNLFESPEEDGPNLTISVSMMLFDTLXGVMTWIEAVPQYQIGPRPW
 YFPTCKYWFPEESDESKHPSGNKRISEIMEEPTHLKLGVSIQLVKVYRGMKV
 AVDLGALVFEGQTFSLGHNGAGKTMISILGTLPFTSGTAVILGKDTRSMSTIR
 QNLGYPQHNVLFWMLIVEEHWYARLGISEKHVAENOMALDVGIPSSKLKSKT
 SOLSGMOKLSVALAFVGGSKVILDEPTAGVDVPSRRGIWELLKLYROGRITILISI
 HMDEADVLSDRIAIIISHGKLCVCGSSFLFKNLQGTGYITLVKKNVESLSSCRNSS
 STVSFLKDEDSVSDAGLQSHEDSDTITDVSAISLNIRKHSSEARLVEDIGHEL
 TIVLPYEAKEGAEVFEHEDDLRLSDGISYGISETTLEEIPKVAESGVDATIS
 DGTLPARNRAFCDKQSCRPFTTEDDAADNDSIDIPESRETDLLSGMDKGSYOVK
 GWKLTQOQFVALLNKRLLIARRSKGFFAQIVIPAVFCIALVFSLVLPFGKVPISLE
 LQPMWYNDSQVLENDAPEDGTLELLNALTQDPGQTRCMEGNPIDETPCQAGEEHW
 TIAVQITIMPLFONGNWMGNPSAQCCSDKTKMLPYCPGAGGLPPQQRKQNTA
 DLDOLLTGRNIDSYKYTYQYIIAASLNKIWNENFYGGFSLGVSNTQALPFSQEVN
 DAXQMKKHLKAKDSADRFLNLSGRFMTGLDTRNNKVVMFNKNGHHAISFLNIVN
 NALLNLRQENKGNESHYGITAFNHPNLTKOQSEVAXMTSYDVLVSICVIAFMGFV
 PASHWFLIQBRVSKAKHLQFISGKQVIVYWLNSFWMDCMYVPAVLVIIIFCQFQ
 KSVYSSTNLPYALLLILLYGWSIPLMPASFEVKIPSTAYVVLISVNLFIGINGSVA
 IFVLELTDLNKLNNIDILKSVFIIIPHCLGRGLDMKNOAMADALEBEGENRVSVA
 PLSDGLGRNLFANAVEGVFELTVLQVREFTPRPVANKLSPINDDEDVDPRREQ
 RILDGCGNDILEKELTKLYRRKXPAVORICVGIPPGCFGLLVNGAGKSTRTGV
 LTGTTVTRGDAPNXXNISLNIEHVQNMGYCQFDATITELLIGREHVEFFALLRGV
 PEKARVGEWAIKELGLVKYKAGIAGNSGNNRKLSTAMALIGPPVYFLDEPTTG
 MDPKRGDGYTVYRIAGSNPDLKPVODFGLAFPGSVXKXKHNMLQVLPSSLSLARI
 PSILSQSKKRLHIEDYSVSQTTLDQVFNFAKQSDGDDHDKDLSLHKNQIVVDVAVILT
 SFLODEKVESYV"
 repeat_region
 29916..30180
 /rpt_family="Alu"
 /rpt_family="Alu"
 31050..31314
 /rpt_family="Alu"
 31692..31973
 /rpt_family="Alu"
 36167..36440
 /rpt_family="Alu"
 37617..37892
 /rpt_family="Alu"
 37999..38166
 /rpt_family="Alu"
 38172..38269
 /rpt_family="Alu"
 38995..39271
 /rpt_family="Alu"
 41547..41565
 /gene="ABCA1"
 /note="A009X28"
 41593..41835
 /rpt_family="Alu"
 41657..41678
 /gene="ABCA1"
 /note="A009X28"
 42320..42584
 /rpt_family="Alu"
 44349..44626
 /rpt_family="Alu"
 48473..48669
 /rpt_family="Alu"
 50547..50614
 /rpt_family="Alu"
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 STS
 repeat_region
 STS
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region


```

exon
/number=18
/40179..140350
/genes="ABCA1"
exon
/number=19
/141340..141471
/genes="ABCA1"
exon
/number=20
/142777..142919
/genes="ABCA1"
exon
/number=21
/143123..143260
/genes="ABCA1"
exon
/number=22
/143963..144183
/genes="ABCA1"
exon
/number=23
/145441..145513
/genes="ABCA1"
exon
/number=24
/146657..146859
/genes="ABCA1"
exon
/number=25
/148527..148575
/genes="ABCA1"
exon
/number=26
/148771..148884
/genes="ABCA1"
exon
/number=27
/150280..150429
/genes="ABCA1"
exon
/number=28
/152078..152202
/genes="ABCA1"
exon
/number=29
/153438..153536
/genes="ABCA1"
exon
/number=30
/156568..156757
/genes="ABCA1"

```

Query Match 94.8%; Score 333.8; DB 9;
Best Local Similarity 98.0%; Pred. No. 5.6e-74;
Matches 350; Conservative 0; Mismatches 2; Indels 5.

QY	1	GGACCTTAGACACCTCTGTACCTCCAC-----CCCCACCCACCCACCTCCGCCAA	55
Db	33665	GGACCTTAGACACCTCTGTACCTCCACCCACCCACCCACCTCCGCCAA	33724
QY	56	CTCCCTAGATGTCGTGGCGGCGCTGAACGTGCGCCCTTTAAGGGGGGGGCCCGGGTCC	115
Db	33725	CTCCCTAGATGTCGTGGCGGCGCTGAACGTGCGCCCTTTAAGGGGGGGGCCCGGGTCC	33784
QY	116	ACGTGCTTTCGTGTGAGTGACTGAACCTACATAACACGAGCGCGGAACGGCGCGGGGAGG	175
Db	33785	ACGTGCTTTCGTGTGAGTGACTGAACCTACATAACGAGCGCGGGAGCGGGCGGGGAGG	33844
QY	176	AGGGAGACACAGGCTTTGACCGATAGTAACTCTCGGCTCGTCAGCCGAAATCTATAA	235
Db	33845	AGGGAGACACAGGCTTTGACCGATAGTAACTCTCGGCTCGTCAGCCGAAATCTATAA	33904
QY	236	AAGGAACCTAGTCCGGCAAAAACCCCGTAATTGCGACGAGAGTGAGTGGGGCCGGGACC	295
Db	33905	AAGGAACCTAGTCTCGCAAAAACCCCGTAATTGCGACGAGAGTGAGTGGGGCCGGGACC	33964
QY	296	CGCAGAGCCGAGCGACCTTCTCTCCCGGGCTGCGCGCAGGGGAGGGCGGGGAGCTC	352
Db	33965	CGCAGAGCCGAGCGACCTTCTCTCCCGGGCTGCGCGCAGGGGAGGGCGGGGAGCTC	34021

RESULT 12	
AC021246/C	
LOCUS	
DEFINITION	

linear RTG 13-JUL-2000
IS SEQUENCE SAMPLING.

ACCESSION	REFERENCE
VERSION	AUTHORS
KEYWORDS	TITLE
SOURCE	JOURNAL
ORGANISM	AUTHORS
	REFERENCE

AC021245
AC021245.2 GI:9A19882
RIG: HTCS-PHASE30.
Homo sapiens.
Homo sapiens.
Homo sapiens.
Eukaryota: Metazoa; Chordata: Vertebrata: Euteleostomi:
Mammalia: Eutheria; Primates: Catarrhini; Hominiidae; Homo.
1 (bases 1 to 69570)
Birren,B., Linton,L., Nusbaum,C., Lander,E.
Homo sapiens chromosome, clone RP11-1N10
Unpublished
2 (bases 1 to 69570)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Becker,T., Beda,F.,
Boguslavsky,L., Bouckgaert,B., Brown,A., Burkett,G., Castle,A.,
Chapel,Y., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,C., Hagos,B., Hesford,A., Horton,L.,
Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lander,T., Lehoczy,J., Levine,R., Liu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGuck,A., McKernan,K.,
McPeatters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivari,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severly,P., Spencer,B., Starge-Thomasan,N.,
Stojanovic,N., Subrananian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

```

* NOTE: This record contains 73 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*

```

1	871:	contig	of 871 bp	in length
872	971:	gap	of 100 bp	
972	1834:	contig	of 863 bp	in length
1835	1334:	gap	of 100 bp	
1935	2804:	contig	of 870 bp	in length
2805	2904:	gap	of 100 bp	
2905	3745:	contig	of 841 bp	in length
3745	3845:	gap	of 100 bp	
3845	4596:	contig	of 851 bp	in length
4697	4796:	gap	of 100 bp	
4797	5540:	contig	of 844 bp	in length
5641	5740:	gap	of 100 bp	
5741	6540:	contig	of 800 bp	in length
6541	6640:	gap	of 100 bp	
6641	7509:	contig	of 869 bp	in length
7510	7609:	gap	of 100 bp	
7610	8479:	contig	of 870 bp	in length


```
* 8480 8579: gap of 100 bp
* 8580 9430: contig of 851 bp in length
* 9431 9530: gap of 100 bp
* 9531 10376: contig of 846 bp in length
* 10377 10476: gap of 100 bp
* 10477 11322: contig of 846 bp in length
* 11323 11422: gap of 100 bp
* 11423 12302: contig of 880 bp in length
* 12303 12402: gap of 100 bp
* 12403 13280: contig of 878 bp in length
* 13281 13380: gap of 100 bp
* 13381 14241: contig of 861 bp in length
* 14242 14341: gap of 100 bp
* 14342 15196: contig of 855 bp in length
* 15197 15296: gap of 100 bp
* 15297 16123: contig of 827 bp in length
* 16124 16223: gap of 100 bp
* 16224 17072: contig of 849 bp in length
* 17073 17172: gap of 100 bp
* 17173 18041: contig of 869 bp in length
* 18042 18141: gap of 100 bp
* 18142 19009: contig of 868 bp in length
* 19010 19109: gap of 100 bp
* 19110 19366: contig of 857 bp in length
* 19367 20066: gap of 100 bp
* 20067 20921: contig of 855 bp in length
* 20922 21021: gap of 100 bp
* 21022 21865: contig of 844 bp in length
* 21866 21965: gap of 100 bp
* 21966 22832: contig of 867 bp in length
* 22833 22932: gap of 100 bp
* 22933 23780: contig of 848 bp in length
* 23781 23880: gap of 100 bp
* 23881 24733: contig of 853 bp in length
* 24734 24833: gap of 100 bp
* 24834 25570: contig of 837 bp in length
* 25571 25770: gap of 100 bp
* 25771 26621: contig of 851 bp in length
* 26622 26721: gap of 100 bp
* 26722 27576: contig of 855 bp in length
* 27577 27676: gap of 100 bp
* 27677 28532: contig of 856 bp in length
* 28533 28632: gap of 100 bp
* 28633 29492: contig of 860 bp in length
* 29493 29592: gap of 100 bp
* 29593 30455: contig of 863 bp in length
* 30456 30555: gap of 100 bp
* 30556 31410: contig of 855 bp in length
* 31411 31510: gap of 100 bp
* 31511 32368: contig of 858 bp in length
* 32369 32468: gap of 100 bp
* 32469 33312: contig of 844 bp in length
* 33313 33412: gap of 100 bp
* 33413 34268: contig of 836 bp in length
* 34269 34368: gap of 100 bp
* 34369 35204: contig of 836 bp in length
* 35205 35304: gap of 100 bp
* 35305 36156: contig of 852 bp in length
* 36157 36256: gap of 100 bp
* 36257 37128: contig of 872 bp in length
* 37129 37228: gap of 100 bp
* 37229 38083: contig of 855 bp in length
* 38084 38183: gap of 100 bp
* 38184 39031: contig of 848 bp in length
* 39032 39131: gap of 100 bp
* 39132 40006: contig of 875 bp in length
* 40007 40106: gap of 100 bp
* 40107 40967: contig of 881 bp in length
* 40968 41067: gap of 100 bp
* 41069 41913: contig of 846 bp in length
* 41914 42013: gap of 100 bp
* 42014 42824: contig of 811 bp in length
* 42825 42924: gap of 100 bp
```

```
* 42925 43776: contig of 852 bp in length
* 43777 43876: gap of 100 bp
* 43877 44752: contig of 876 bp in length
* 44753 44852: gap of 100 bp
* 44853 45724: contig of 872 bp in length
* 45725 45824: gap of 100 bp
* 45825 46643: contig of 819 bp in length
* 46644 46743: gap of 100 bp
* 46744 47599: contig of 856 bp in length
* 47600 47699: gap of 100 bp
* 47700 48551: contig of 852 bp in length
* 48552 48651: gap of 100 bp
* 48652 49485: contig of 834 bp in length
* 49486 49585: gap of 100 bp
* 49586 50440: contig of 855 bp in length
* 50441 50540: gap of 100 bp
* 50541 51404: contig of 864 bp in length
* 51405 51504: gap of 100 bp
* 51505 52372: contig of 868 bp in length
* 52373 52472: gap of 100 bp
* 52473 53328: contig of 856 bp in length
* 53329 53428: gap of 100 bp
* 53429 54268: contig of 840 bp in length
* 54269 54368: gap of 100 bp
* 54369 55229: contig of 861 bp in length
* 55230 55329: gap of 100 bp
* 55330 56197: contig of 868 bp in length
* 56198 56297: gap of 100 bp
* 56298 57163: contig of 866 bp in length
* 57164 57263: gap of 100 bp
* 57264 58130: contig of 867 bp in length
* 58131 58230: gap of 100 bp
* 58231 59082: contig of 852 bp in length
* 59083 59182: gap of 100 bp
* 59183 60020: contig of 838 bp in length
* 60021 60120: gap of 100 bp
* 60121 60983: contig of 863 bp in length
* 60984 61083: gap of 100 bp
* 61084 61935: contig of 852 bp in length
* 61936 62035: gap of 100 bp
* 62036 62866: contig of 831 bp in length
* 62867 62966: gap of 100 bp
* 62967 63827: contig of 861 bp in length
* 63828 63927: gap of 100 bp
* 63928 64783: contig of 856 bp in length
* 64784 64883: gap of 100 bp
* 64884 65740: contig of 857 bp in length
* 65741 65840: gap of 100 bp
* 65841 66684: contig of 844 bp in length
* 66685 66784: gap of 100 bp
* 66785 67651: contig of 867 bp in length
* 67652 67751: gap of 100 bp
```

Query Match

Best Local Similarity 77.2%; Score 271.6; DB 2; Length 69570;

Matches 289; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
QY 31 CCCACCCACCCACCTCCCCCAACTCCCTAGATGTCTCGTGGCGGCTGAACGTGGCC 90
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 67642 CCCCCNNCCCCACCACCTCCCCCAANTATAGATGTCTCGTGGCGGCTGAACGTGGCC 67583

QY 91 COTTAAAGGGGGGGGGGGGGTCCACGTGCTTCTGCTGAGTGACTGAACATACATAAC 150
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 67582 CGTAAAGGGGGGGGGGGGGTACACGTGCTTCTGCTGAGTGACTGAACATACAC 67523

QY 151 AGAGCCCGGAACGGGGGGGGGAGGAGGACAGACAGGCTTTGACCGATAGTACCTCT 210
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 67522 AGAGCCCGGAACGGGGGGGGGAGGAGGAGGACAGACAGGCTTTGCTGATAGTACTTCT 67463

QY 211 GGGTCGGTGAGCGGAATCTATATAAGAACTAGTCCCGGCAAAACCCGTAATCGG 270
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 67462 GCGCTCGGTGAGCGGAATCTATATAAGAACTAGTCCCGGCAAAACCTCTGTAATCGG 67403
```

```
QY 271 AGCGAGTGTAGTGGCGGGGACCCGAGCGAGCGAGCCCTTCTCTCCGGGCTGC 330
|||||
Db 67402 ACGAGAGTGTAGTGGCGGGGACCCGAGCGAGCGAGCCCTTCTCTCCGGTCTCC 67343
QY 331 GCGAGGCGAGGCGGGGAGCTC 352
|||||
Db 67342 GACAGGACAGGCGGGGAGCTC 67321

RESULT 13
AX351031 AX351031 2893 bp DNA linear PAT 06-FEB-2002
LOCUS Sequence 3 from Patent WO0183746.
DEFINITION
ACCESSION AX351031
VERSION AX351031.1 GI:18616387
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE
AUTHORS Rosier-Montus M.F., Prades C., Lemoine C., Naudin L., Deneffe P.,
Brewer B., Duvergier N., Remaley A. and Santamarina-Fojo S.
TITLE Regulatory nucleic acid sequences of the abcl gene
JOURNAL Patent: WO 0183746-A 3 08-NOV-2001;
Aventis Pharma S.A. (FR)
FEATURES
Location/Qualifiers
Source 1..2893
/organism="Homo sapiens"
/ds_xref="taxon:9606"
BASE COUNT 745 a 681 c 769 g 698 t
ORIGIN

Query Match 69.4%; Score 244.4; DB 6; Length 2893;
Best Local Similarity 97.7%; Pred. No. 2.8e-51;
Matches 260; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

QY 1 GGACCTTAGACACCTGTCTACCTCCAC-----CCCCACCCACCCACCTCCCCCAA 55
|||||
Db 2628 GGACCTTAGACACCTGTCTACCTCCACCCACCCACCCACCTCCCCCAA 2687
QY 56 CTCCTAGATGTGTCGTGGGGCGGTGAACGTGCGCGGTTAAGGGCGGGCGCGCTCC 115
|||||
Db 2688 CTCCTAGATGTGTCGTGGGGCGGTGAACGTGCGCGGTTAAGGGCGGGCGCGCTCC 2747
QY 116 ACCTGCTTTCTGCTGACTGAAGTACATAACAGAGCGCGGAGCGGGGAGG 175
|||||
Db 2748 ACCTGCTTTCTGCTGACTGAAGTACATAACAGAGCGCGGAGCGGGGAGG 2807
QY 176 AGGAGACACAGGCTTTCACCGATAGTAACCTCTCGCTCGGTGAGCGGATCTATAA 235
|||||
Db 2808 AGGAGACACAGGCTTTCACCGATAGTAACCTCTCGCTCGGTGAGCGGATCTATAA 2867
QY 236 AGGAACTAGTCCCGCAAAACCCC 261
|||||
Db 2868 AGGAACTAGTCCCGCAAAACCCC 2893

RESULT 14
AC021345/c AC021345 90698 bp DNA linear HTG 13-JUL-2000
LOCUS Homo sapiens clone RP11-24J9, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
ACCESSION AC021345
VERSION AC021345.2 GI:9130845
KEYWORDS HTG; HTGS_PHASEC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 90698)
REFERENCE Birren B., Linton L., Nusbaum C. and Lander E.
AUTHORS Homo sapiens, clone RP11-24J9
TITLE Unpublished
JOURNAL
```

REFERENCE
AUTHORS

2 (bases 1 to 90698)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Bedalov, F., Boguslavsky, L., Borkigaiter, B., Brown, A., Burkett, G., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cocke, P., DeArellano, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J., Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kahn, L., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meid, J., Menes, L., Morrow, J., Naylor, J., Norman, C. E., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tizrell, A., Vassiliev, E., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zocay, M.

Direct Submission

Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced 91:6705761.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIER

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4483

Center clone name: 24_J-9

* NOTE: This record contains 92 individual
* sequencing reads that have not been assembled into
* contigs. Rups of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 910: contig of 910 bp in length
911 1010: gap of 100 bp
1011 1873: contig of 863 bp in length
1874 1973: gap of 100 bp
1974 2824: contig of 851 bp in length
2825 2924: gap of 100 bp
2925 3902: contig of 878 bp in length
3903 3902: gap of 100 bp
3903 4816: contig of 914 bp in length
4817 4916: gap of 100 bp
4917 5759: contig of 843 bp in length
5760 5859: gap of 100 bp
5860 6764: contig of 905 bp in length
6765 6864: gap of 100 bp
6865 7747: contig of 883 bp in length
7748 7847: gap of 100 bp
7848 8755: contig of 908 bp in length
8756 8855: gap of 100 bp
8856 9753: contig of 898 bp in length
9754 9853: gap of 100 bp
9854 10757: contig of 904 bp in length
10758 10857: gap of 100 bp
10858 11732: contig of 875 bp in length
11733 11832: gap of 100 bp
11833 12739: contig of 907 bp in length
12740 12840: gap of 100 bp
12840 13710: contig of 871 bp in length
13711 13810: gap of 100 bp

13811	14664:	contig of 874 bp	in length
14885	14784:	gap of 100 bp	
14785	15662:	contig of 878 bp	in length
15663	15762:	gap of 100 bp	
15763	16677:	contig of 915 bp	in length
16678	16777:	gap of 100 bp	
16779	17678:	contig of 901 bp	in length
17679	17778:	gap of 100 bp	
17779	18679:	contig of 901 bp	in length
18680	18779:	gap of 100 bp	
18780	19632:	contig of 853 bp	in length
19633	19732:	gap of 100 bp	
19733	20634:	contig of 902 bp	in length
20635	20734:	gap of 100 bp	
20735	21620:	contig of 886 bp	in length
21621	21720:	gap of 100 bp	
21721	22575:	contig of 859 bp	in length
22580	22679:	gap of 100 bp	
22680	23588:	contig of 889 bp	in length
23569	23668:	gap of 100 bp	
23669	24554:	contig of 886 bp	in length
24555	24554:	gap of 100 bp	
24555	25591:	contig of 867 bp	in length
25522	25621:	gap of 100 bp	
25622	26487:	contig of 866 bp	in length
26488	26587:	gap of 100 bp	
26588	27464:	contig of 877 bp	in length
27465	27564:	gap of 100 bp	
27565	28466:	contig of 902 bp	in length
28467	28566:	gap of 100 bp	
28567	29464:	contig of 898 bp	in length
29465	29564:	gap of 100 bp	
29565	30447:	contig of 883 bp	in length
30448	30547:	gap of 100 bp	
30548	31453:	contig of 906 bp	in length
31454	31553:	gap of 100 bp	
31554	32452:	contig of 899 bp	in length
32453	32552:	gap of 100 bp	
32553	33447:	contig of 895 bp	in length
33448	33547:	gap of 100 bp	
33548	34435:	contig of 888 bp	in length
34436	34535:	gap of 100 bp	
34536	35433:	contig of 898 bp	in length
35434	35533:	gap of 100 bp	
35534	36440:	contig of 907 bp	in length
36441	36540:	gap of 100 bp	
36541	37422:	contig of 882 bp	in length
37423	37522:	gap of 100 bp	
37523	38402:	contig of 880 bp	in length
38403	39622:	gap of 100 bp	
39603	39460:	contig of 878 bp	in length
39480	40407:	contig of 927 bp	in length
40408	40507:	gap of 100 bp	
40508	41405:	contig of 898 bp	in length
41406	41505:	gap of 100 bp	
41506	42399:	contig of 894 bp	in length
42400	42499:	gap of 100 bp	
42500	43357:	contig of 858 bp	in length
43358	43457:	gap of 100 bp	
43458	44356:	contig of 899 bp	in length
44357	44456:	gap of 100 bp	
44457	45325:	contig of 869 bp	in length
45326	45425:	gap of 100 bp	
45426	46305:	contig of 880 bp	in length
46306	46405:	gap of 100 bp	
46406	47302:	contig of 897 bp	in length
47303	47402:	gap of 100 bp	
47403	48293:	contig of 891 bp	in length
48294	48393:	gap of 100 bp	
48394	49280:	contig of 887 bp	in length
49281	49380:	gap of 100 bp	
49381	50350:	contig of 870 bp	in length

QY	48	CCGCCCACTCCCTAGATGTGCTGGGGGCTGAACCTGCCCGTTTAAGGCGCGGCC	107
Db	87745	CCCCACTCCCCACATAGTGTGCTGGCGCTGAACCTGCCCGTTTAAGGCGCGGCC	87687
QY	108	CCGGCTCCACGTGCTTCTGCTGAGTACTGAACCTACATAACAGAGCCGGGAACCGGG	167
Db	87686	CCGGCTCCACGTGCTTCTGCTGAGTACTGAACCTACATAACAGAGCCGGGAACCGGG	87628
QY	168	CGGGAGGAGGAGAGACAGAGCTTTGACCGATAGTAACTCTCGGTGCGAGCCGA	227
Db	87627	CGGGAGGAGGAGAGACAGAGCTTTGACCGATAGTAACTCTCGGTGCGAGCCGA	87568
QY	226	ATCTATRAAGGAACCTAGTCCCGGCAAAACCCGTAATCCGAGCAGAGTGAAGTGGG	287
Db	87567	ATCTATRAAGGAACCTAGTCCCGGCAAAACCCGTAATCCGAGCAGAGTGAAGTGGG	87509
QY	288	CCGGGACCGGACAGCCGAGGCGGACCTCTCTCTCCGGGCTGCGGACAGGCGCGGG	347
Db	87508	CCGGGACCGGACAGCCGAGGCGGACCTCTCTCTCCGGGCTGCGGACAGGCGCGGG	87449
QY	348	AGCTC	352
Db	87448	AGCTC	87444

Query Match: 64.5%; Score 227; DB 2; Length 90698;

Best Local Similarity 90.2%; Pred. No. 4.7e-47;

Matches 275; Conservative 0; Mismatches 27; Indels 3; Gaps 3;

RESULT 15

AD258627

LOCUS AF258627 697 bp mRNA linear PRI 11-MAY-2000
 DEFINITION Homo sapiens ATP binding cassette transporter 1 (ABCA1) mRNA, partial cds.
 ACCESSION AF258627
 VERSION AF258627.1 GI:7769707
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 697)
 AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C., Aouizerat,B.E., Fielding,C.J., and Kane,J.P.
 TITLE Analysis of ABCA1 gene 5' end: additional peptide sequence, promoter region, and four polymorphisms
 JOURNAL Biochem. Biophys. Res. Commun. 271 (2000) In press
 REFERENCE 2 (bases 1 to 697)
 AUTHORS Pullinger,C.R., Hakamata,H., Duchateau,P.N., Eng,C., Aouizerat,B.E., Fielding,C.J., and Kane,J.P.
 TITLE Direct Submission
 JOURNAL Submitted (19-APR-2000) Cardiovascular Research Institute, University of California, San Francisco, 505 Parnassus Avenue, San Francisco, CA 94143-0130, USA
 FEATURES
 Location/Qualifiers
 1..697
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="9"
 /map="9q31"
 /tissue_type="placenta"
 1..>697
 /gene="ABCA1"
 /note="ABCA1"
 396..>697
 /gene="ABCA1"
 /note="membrane-bound"
 /codon_start=1
 /product="ATP binding cassette transporter 1"
 /protein_id="AAF69513.1"
 /db_xref="GI:7769708"
 /translation="MACWPOLRLIMKJFERRQTCOLLLEVAWPEFIFLLISVRL
 SYPPYEQHECHFPNKAMPSAGTIPWVQGLICNANNPCFRYPTEAPGVVGNFNKS"

BASE COUNT 152 a 198 c 190 g 156 t 1 others
 ORIGIN

Query Match 49.1%; Score 173; DB 9; Length 697;
 Best Local Similarity 100.0%; Pred. No. 3.1e-33;
 Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 180 AGAGCACAGGCTTTGACCGATAGTAACTCTGCGCTGGTGCAGCCGCAATCTATAAAGG 239
 Db 1 AGAGCACAGGCTTTGACCGATAGTAACTCTGCGCTGGTGCAGCCGCAATCTATAAAGG 60
 QY 240 AACTAGTCCCGGCAAAACCCGTAATTGCGAGGAGAGTCAAGTGGGCGGCGGACCGCA 299
 Db 61 AACTAGTCCCGGCAAAACCCGTAATTGCGAGGAGAGTCAAGTGGGCGGCGGACCGCA 120
 QY 300 GAGCCGAGCGAGCCCTCTCTCCGGGCTGGGCGAGGCGGCGGAGCTC 352
 Db 121 GAGCCGAGCGAGCCCTCTCTCCGGGCTGGGCGAGGCGGCGGAGCTC 173

Search completed: February 4, 2003, 01:32:23
 Job time : 1336.79 secs